THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ENTITLED:

'IMPROVING THE MEASUREMENT OF VISUAL ACUITY IN CLINICAL PRACTICE AND CLINICAL RESEARCH'

DANIEL ASHLEY ROSSER

INSTITUTE OF OPHTHALMOLOGY

UNIVERSITY COLLEGE LONDON

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

I would like to thank my supervisor Ian Murdoch for his help and support throughout this project. His academic council has been of the highest calibre and I have learned an immeasurable amount from him. More than this however, I am grateful for his perpetually cheerful demeanor, and in particular his relentless and infectious enthusiasm without which this thesis would never have been completed.

I would like to thank Alistair Laidlaw for providing the inspiration for this work. He has remained a committed devotee of this project throughout, and his input has been invaluable.

I would like to thank Simon Cousens for his contribution to this work. Over and above his statistical advise, his unique insight into the subject provided the inspiration for the final two experiments described herein.

I would like to thank Louise Culham for her support and advice.

I would like to thank the staff of the Department of Epidemiology and International Eye Health at the Institute of Ophthalmology, and in particular Helen Baker and Pak Sang Lee (MPhil) for their help and support.

I would like to thank Winnie Nolan, Rupert Bourne and Brendan Dineen for allowing their data to appear in Appendix B.

Last but by no means least, I would like to thank my wife Nichola for putting up with everything and generally being wonderful.

ABSTRACT

This thesis considers the relationship between the design of a visual acuity test and various aspects of its performance. Using contemporary test design theory, novel tests are developed and evaluated in an attempt to better meet the requirements of a visual acuity test most pertinent to clinical practice, clinical research, and population based surveys.

The acuity test of choice in clinical practice is the Snellen chart, a test whose usefulness is limited by several design flaws. Clinical researchers favour the ETDRS logMAR chart which employs robust design principles, but is time consuming to use. A chart featuring an abbreviated ETDRS design was developed and its performance compared with that of the ETDRS and Snellen charts. The prototype chart allows acuities to be measured in half the time of the ETDRS chart with greater precision than the Snellen chart. A tumbling-E version of this chart has been successfully employed in population based surveys in Thailand, Bangladesh and Mongolia.

It was noted during the study that the precision of even ETDRS acuities was relatively poor. A computerised version of the ETDRS test was developed and used to investigate the repeating and averaging of acuities as a means to improve precision. Whilst prolonging test time, the computerised test allowed acuities to be measured with improved precision.

Optical defocus was investigated as a potential source of reduced precision in visual acuity testing. It was shown that even small degrees of defocus may significantly reduce test precision. An approach which considers test performance in terms of sensitivity and specificity was developed. A mathematical model was used to show that current methods of using estimates of precision to identify clinically important change, are overly optimistic. Predictions derived using the model were shown to agree well with empirical findings.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Number	Title
Figure 16.3.1	Predicted sensitivity of repeated acuity measurements to detect changes in visual acuity

1. OVERVIEW

Visual acuity is the most important measure of visual function both in the clinical environment and within clinical research. It is a measure of a subject's ability to resolve detail. As well as determining the degree of detail a subject can resolve, visual acuity tests are also used to detect changes in their ability to resolve detail e.g. which may be the result of a disease process or treatment. Visual acuity is normally measured by asking the subject to attempt a series of progressively smaller letter-stimuli on a chart until the subject is incapable of recognising the stimulus. The level of visual acuity is then specified according to the angular subtense of the smallest recognisable letter. If visual acuity is to be specified in this way, then the visual acuity chart should be designed such that at each stimulus size, the task is identical apart from the fact that the stimuli size varies. As well as stimulus size being the sole determinant of difficulty, the measurement scale should be sufficiently fine to be able to detect changes in resolving ability which are of clinical importance, and should cover the range of values which may be encountered.

In clinical research, visual acuity is typically measured using contemporary charts whose design would appear to satisfy the aforementioned requirements. However, the most common test of visual acuity in clinical practice is the Snellen chart whose design introduces a number of variables which have the potential to confound the variable of interest; letter size. In addition to this, the scale is coarse, and is truncated at the high acuity end. Why then have contemporary charts not been adopted into routine clinical practice? The author believes that there may be a number of reasons for this:

- contemporary charts may be perceived as too cumbersome for clinical practice
- there may be poor understanding of the potential benefits of contemporary charts
- there may be reluctance to use an unfamiliar notation
- there may be an overoptimistic impression of the usefulness of the Snellen chart (due in part to the lack of rigorous measurement procedures).

The first of a series of studies investigated the hypothesis that an abbreviated version of the gold standard ETDRS logMAR chart allow visual acuity to be measured quickly

enough for clinical practice, but with greater precision than the Snellen chart. The results demonstrated that a reduced logMAR (or RLM) chart can measure acuities in half the time of the ETDRS chart, and with greater precision than the Snellen chart. The study also demonstrated the trade off which exists between test precision (which determines the ability to detect change) and test duration, whereby a finer scale increment improves precision but lengthens the test due to the increased total number of stimuli. A 'tumbling E' version of one of the prototype charts was produced and has since been used in two large prevalence surveys and a randomised control trial (based in Thailand, Bangladesh and Mongolia respectively).

The precision of even the ETDRS chart in the initial study was such that the ability of the test to detect change appeared limited. A study was designed to determine whether reducing the size of the effective scale increment through averaging repeated acuity thresholds could offer improved precision and hence detect clinical change earlier. A computerised test (the 'PC-test') was designed to display the acuity stimuli as well as average the resultant acuity scores. The results showed that acuities measured using such a system were accurate compared with those measured using the ETDRS chart as well as being more precise. A separate study compared the performance of the ETDRS chart and the PC-test in amblyopic subjects. The results of this study again demonstrated that acuities measured using the PC-test are more precise than ETDRS acuities, and hence able to detect changes earlier.

It was noted that despite using the same methodology and examiner, the estimates of ETDRS precision from the first three studies varied considerably. The same can be said of published estimates of precision for the ETDRS chart (even after accounting for the effect of differing scoring methods). Following a review of such published studies, uncorrected refractive error was identified as the most likely confounding variable. A study was carried out to investigate the effect of optical defocus upon the precision of the ETDRS chart. The results of this study showed that even degrees of optical defocus as small as +0.50 dioptres can reduce the precision of ETDRS acuities considerably, with +1.00 dioptres of blur exerting an even larger effect.

This thesis has followed convention by specifying the precision of an acuity test in terms of the width of the 95% test-retest range (TRR). The width of the 95% TRR has

been widely advocated as a cut-off (or 'change-criterion') by which to decide whether or not measured changes in acuity can be attributed to a true change in clinical status. The author has proposed that a clinician using the 95% TRR in this way is ensuring a specificity of 95%, but remains unaware of the test's level of sensitivity to change. A study was conducted in which varying degrees of acuity change were simulated by manipulating the chart viewing distance. Specificity and sensitivity to change were estimated for ETDRS and PC-test acuities (when using the width of the 95% TRR as a change-criterion). The results showed that the sensitivity of ETDRS acuities to changes of equal magnitude to the 95% TRR, is approximately 50%. Changes in visual acuity may only be detected with sensitivity of 95% or more when their magnitude is around twice that of the 95% TRR. The performance of the PC-test was superior to that of the ETDRS chart either when compared using the 95% TRR, or using receiver operator characteristic curves. Finally, a simple statistical model was developed to predict the sensitivity to a given degree of change from an estimate of the test's 95% TRR. Predictions made using this model agreed well with the empirical findings from the previous study.

2. INTRODUCTION

2.1. BACKGROUND

Man's interaction with his environment has been greatly facilitated by the richness of his perception of that environment. The body's visual system has developed a degree of complexity commensurate with that of the visual world within which it functions. There are numerous facets of visual function which are seamlessly combined to create a rich visual perception. These facets vary from the very simple, such as the ability to detect the presence of a spot of light against a dark background, through to more complex functions such as the perception of depth through binocular image disparity, or the recognition of familiar objects, independent of orientation and lighting. Vision scientists are familiar with the process of isolating aspects of visual function such that their contribution to our overall perception can be elucidated. For the clinician, and the clinical researcher, there has traditionally been one aspect of visual function which has taken precedent over the others: the ability to resolve detail. As Westheimer[1] put it, '...resolution of detail constitutes the essence of eyes and vision'. There may be several reasons for this predominance. The ability of the visual system to resolve detail is highly dependant upon both the performance of the eye's optical system, and the spatial arrangement of the light receptor cells of the central retina. Many of the more common visually significant ocular abnormalities affect the ocular media or the retina. As such, a measure of the ability to resolve detail would be acutely sensitive to a range of abnormalities. From the patient's perspective, those eye conditions which cause significant visual symptoms often do so through their capacity to affect the ability to resolve detail, rather than other aspects of visual function e.g. visual field/colour vision. Another reason for the predominance of the ability to resolve detail in the assessment of visual function may be the dramatic degree to which it is influenced by refractive error, which is relatively prevalent.

2.2. Defining Visual Acuity

The term 'visual acuity' was coined by Donders[2] circa 1860 to describe 'the sharpness of vision'. He proposed that it be defined as the ratio between a subject's performance and a standard performance, a notion which influenced the form of notation used in the first clinical test of visual acuity, the Snellen chart. More recently, Bailey[3] has defined visual acuity as 'the spatial resolving capacity of the visual system', in other words a measure of the degree of detail which can just be discerned by the observer. We can relate the 'detail' in an image to those areas which feature rapid changes of luminance across the image's spatial profile; i.e. sharp edges. Since the suggestion that the visual cortex processes spatial frequencies rather than particular features of an image[4], images have been considered as being complex waveforms which may be broken down in to a number of component sine-waves of varying spatial frequencies. The sharp edges in an image contain high spatial frequencies. Hence, the level of visual acuity relates to the ability to resolve high spatial frequencies. The ability of the visual system to resolve high spatial frequencies is maximum at high contrast, and as such the term visual acuity generally refers to the spatial resolving power of the eye at high contrast.

2.3. Visual Acuity – One Aspect of Visual Discrimination

The various facets of visual discrimination can be divided into three broad categories: light discrimination, spatial discrimination, and temporal discrimination. Each category can be further subdivided as shown in Fig 2.3.1. Although each element of discrimination has been closely studied, not all have proved of equal importance from a clinical perspective.

2.3.1. *Light discrimination*

Brightness sensitivity – this represents the ability to detect a dim light against a dark background. A variant of the brightness sensitivity test is carried out clinically in the form of perimetry. Such tests are common in the assessment of conditions which affect peripheral vision, and may be used to monitor disease or to assess visual function for example with respect to safety to drive. As the peripheral retina is not specialised for spatial resolution, a brightness sensitivity test is more useful in the assessment of conditions such as glaucoma and neurological conditions affecting the visual pathway. Most perimeters carry out estimates of eccentric brightness sensitivity threshold at various locations in a subject's peripheral vision. The results of these thresholds are then pooled to provide an overview of peripheral visual function. Although such tests could be carried out against a dark background, they typically depart slightly from the definition of brightness sensitivity given above in that for practical reasons they generally employ photopic levels of background illumination rather than a dark background[5]. Peripheral brightness sensitivity can be thought of as complimentary to visual acuity and in combination these two tests provide an excellent overview of a visual function.

Brightness discrimination – this is the ability to detect threshold changes in brightness or differences in the luminance of light sources. Contrast sensitivity testing is a form of brightness discrimination which is frequently performed in clinical practice (see section 2.5.6).

Colour discrimination – this is the ability to differentiate between different colours. Defects in colour discrimination may be congenital or acquired. Tests of colour discrimination have considerable importance from both a clinical and occupational

standpoint. As the same retinal receptors are responsible for both, there will be a variety of retinal conditions which affect both colour discrimination and visual acuity. The tests are however complimentary as although visual acuity is the primary test of central retinal function, conditions exist (such as some toxic and neuro-ophthalmic abnormalities) in which colour vision tests may be more sensitive than visual acuity.

2.3.2. *Spatial discrimination*

Visual acuity – the various classifications of acuity tests are dealt with in detail in section 2.5.

Distance discrimination – this represents the ability to estimate the distance of an object or discriminate between the relative distances of multiple objects. The visual system uses many monocular cues to help judge depth, but in its purest form depth perception relates to the judgement of relative distance through the higher processing of binocular image disparity. The ability to judge depth in the absence of monocular cues is termed stereopsis. Stereopsis is acutely sensitive to defects in the development of binocular visual function, therefore tests of stereopsis are commonplace in the clinical assessment of visual function in children. Tests of stereopsis can be considered complimentary to visual acuity tests as although the detection of image disparity requires good monocular acuity in each eye, stereopsis may be absent even in the presence of good monocular acuities. Stereopsis is therefore specific to binocular visual function.

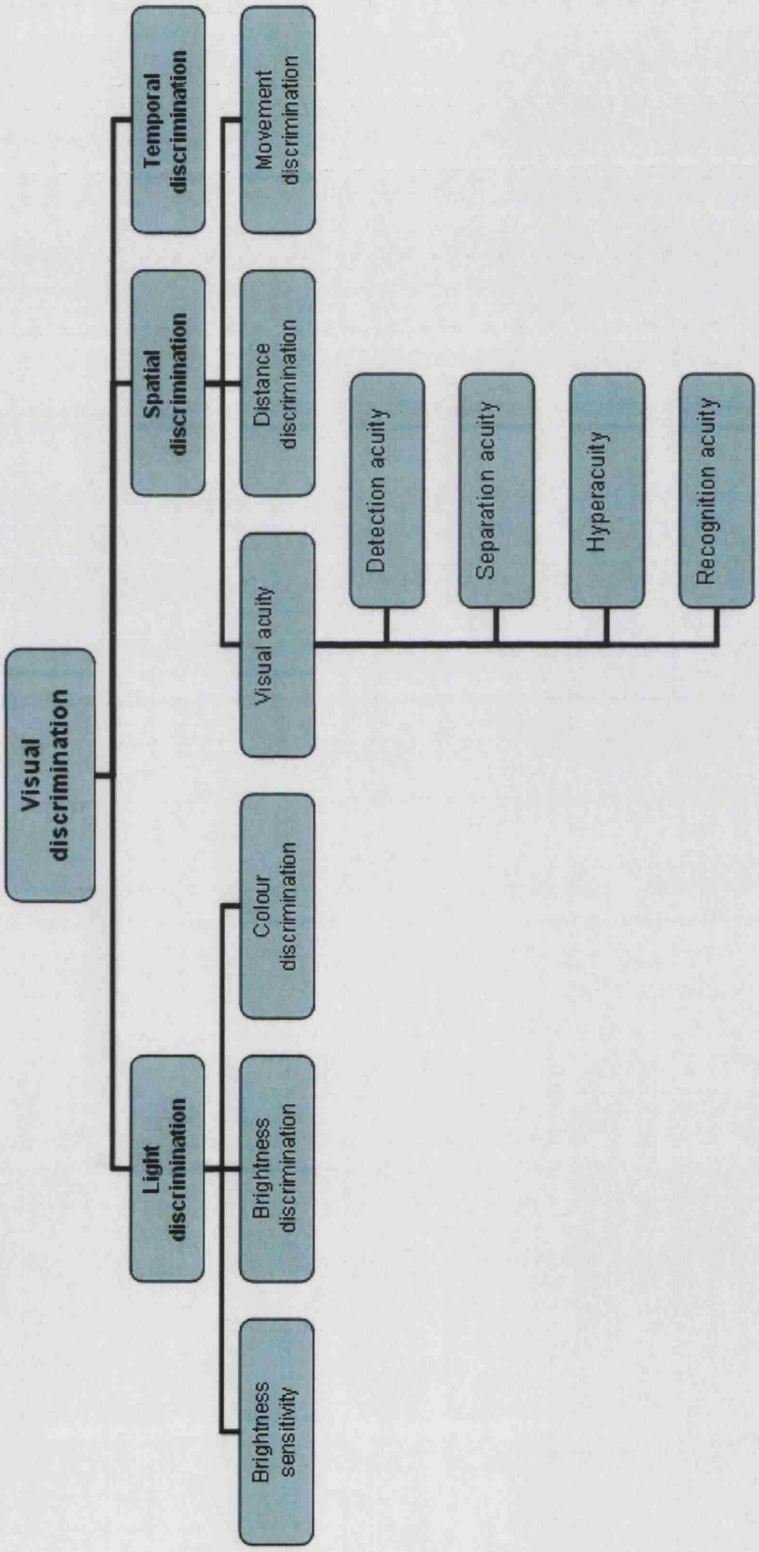
Movement discrimination – this is the ability to detect relative or absolute angular motion. Although tests of motion detection are not commonplace, motion detection may be preferentially affected by glaucoma[6 7] and other neuropathies affecting the visual pathway[8].

2.3.3. *Temporal discrimination*

This relates to the growth and decay of sensations caused by time-varying stimuli. As for movement discrimination, tests of temporal discrimination are not commonplace, but there is again evidence for some selective deficits in certain conditions. For

example there is evidence of an elevated critical flicker fusion threshold in some neurological conditions[9 10].

Figure 2.3.1. Diagrammatic representation of the components of visual discrimination



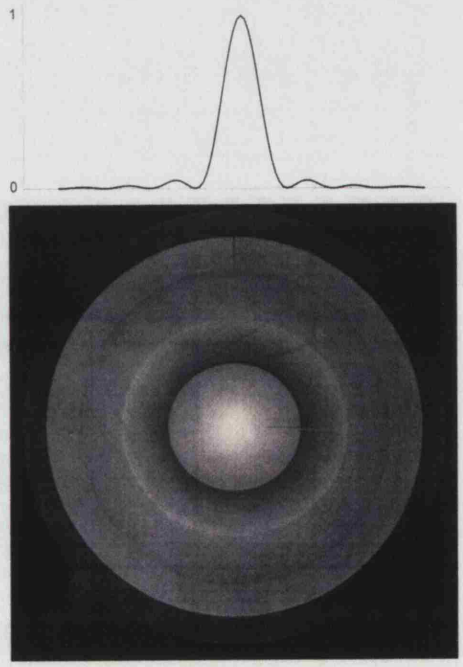
2.4. Physiological Basis of Visual Acuity

Before examining the various types of visual acuity test, it is useful to consider the physiological basis of visual acuity, as a knowledge of this facilitates a critical appraisal of the merits of various tests for the assessment of spatial resolving power. When presented with a stimulus, the visual system's perception of the stimulus is dependent upon a series of optical and neural processes. In the case of visual acuity, we are concerned with assessing the upper limit of spatial resolution. Limiting factors in this process with respect to spatial resolution may be potentially optical and/or neural[1 11].

2.4.1. *Optical factors*

Even in the absence of refractive error, a point object is not imaged as a point on the retina but as an annulus of light (the 'Airy disk') surrounded by feint rings. This distribution of light on the retina from a point object is called the point spread function (see Fig 2.4.1).

Figure 2.4.1. The point spread function of the human eye.



This occurs due to diffraction of light at the pupil margin as it enters the eye and therefore the diameter of the central annulus of light varies with pupil size. The relationship is as follows:

$$\omega = 2.44 \lambda / p$$

where ω is the diameter of the Airy disk (radians), λ is the wavelength of light and p is the pupil size. The inverse relationship between the diameter of the Airy disc and pupil size suggests that the highest degree of resolution will be achieved at maximum pupil size. The fact that this is not the case is explained by the presence of various aberrations within the optical system of the eye. Along with diffraction effects, monochromatic and chromatic aberrations degrade the quality of the retinal image but the effect of these aberrations increases with increasing pupil size. It has been shown that there is an optimum pupil size of around 2.4mm at which the combined effects of diffraction and aberration upon the diameter of the point spread function are at a minimum[12-14]. According to the 'Raleigh criterion', the images of two point objects formed by an optical system, can just be resolved as separate when the centre of one Airy disc coincides with the edge of the other (see section 2.5.2 and Fig 2.5.1). Hence for an optimal pupil size of 2.4mm, the optical limit on resolution is slightly better than 1 minute of arc (equivalent to Snellen 6/6).

2.4.2. *Neural factors*

The neural limit to resolution is dictated by the retinal receptor density and the degree of neural convergence between these cells and the visual cortex. At the fovea, the centres of adjacent cones are separated by around 2μm and there is a 1:1 relationship between photoreceptors and ganglion cells i.e. zero convergence. If we assume that two spots of light can just be resolved as being separate when they fall on the centres of two cones separated by a single unstimulated cone, then at the fovea, the limit of resolution occurs at an image separation of 4μm. This represents an angular subtense of just under 1 minute of arc.

The optical and neural limitations on visual acuity are therefore similar[3 14 15]. It is interesting to consider this in the light of recent attempts to achieve 'supernormal' levels of visual acuity through the use of 'wavefront guided' laser refractive surgery

techniques[16-18]. This process involves measuring the eye's higher order aberrations using abberometry such that the corneal ablation can be 'customised' in an attempt to eliminate these higher order aberrations along with the ametropia. The oft cited data of Campbell and Green[19] have been reviewed by Applegate in an attempt to decide whether such 'custom ablations' can offer supernormal acuity[20]. Applegate has suggested that for those individuals whose best corrected acuity falls short of Snellen 6/3, the factors limiting acuity are likely to be optical. Hence for this group, correction of higher order aberrations may offer improved acuity up to a limit of around 6/3 at which point neural factors are likely to prevent any further improvement. Those who already enjoy a best corrected acuity of 6/3, are unlikely to achieve finer resolution, although even for this group, correction of higher aberrations may offer improved image contrast. For non-foveal vision, the reduced photoreceptor density along with increased neural convergence dictates that the factors limiting peripheral acuity are neural rather than optical.

It is theoretically possible that the involuntary eye movements to which the eyes are constantly subject may improve resolving power through enhancing the differential activity on which visual acuity depends. In fact, some theories of visual acuity have been based upon the assumption that retinal receptors must scan a contour to maximise the degree of spatial discrimination[21 22]. Such theories appear to have been disproved by the discovery that visual acuity remains unchanged when the effect of eye movements is removed through retina image stabilisation[23]. Indeed, a review of research into eye movements and visual acuity has suggested that such temporal factors are of minor consequence in comparison with the importance of spatial factors in determining visual acuity[24]. Involuntary eye movements are however beneficial in avoiding transient visual loss during attempted steady fixation of a visual acuity stimulus[25].

2.5. METHODS OF MEASUREMENT

The methods of measuring visual acuity have historically been divided into various categories according to the type of resolving task employed[15 24 26-29]. In general, authors have used similar classifications: the following is after Bailey[3] and Riggs[24].

2.5.1. *Detection acuity*

This represents the ability to detect the presence of a target against a background without regard to its form. This method is employed in some tests for measuring children's vision[30 31]. As a method of assessing the spatial characteristics of the visual system, it has limitations. These limitations are most obvious for the case of a circular spot of light viewed against a dark background (although the same argument applies to the detection of a linear target). For large spot sizes the width of the retinal image is determined in the main by the size of the spot. However as the spot size decreases, the luminance distribution of the retinal image tends towards the point spread function the shape of which is determined by diffraction effects (see section 2.4.1). As the spot size decreases below about 1 minute of arc, the shape of the retinal distribution remains unaltered but its amplitude reduces. As a result the test becomes one of contrast discrimination rather than one of spatial resolving capacity. Also, for spot sizes below 1 minute of arc, the reduced amplitude of the retinal image distribution which accompanies reduced spot size can be compensated for by increasing the intensity of the object. Therefore, providing the object is sufficiently bright, we can detect objects with an infinitely small angular subtense (as in the example of our ability to easily perceive distant stars[15]).

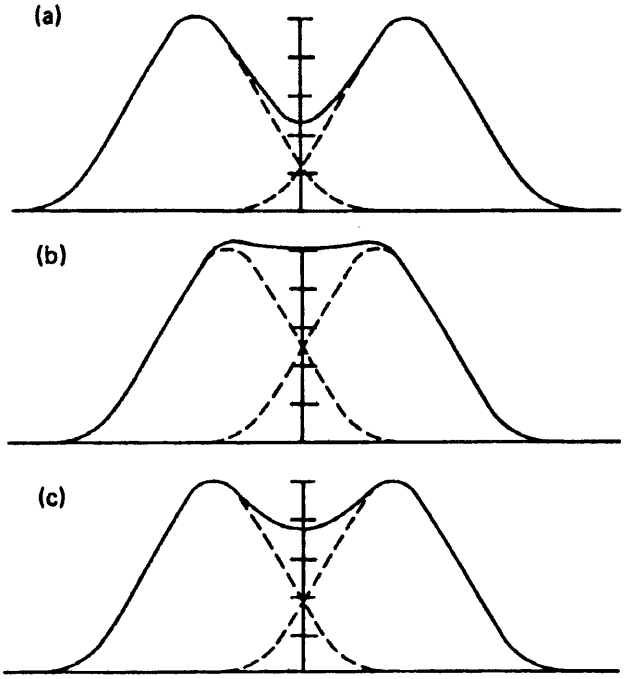
The case of a dark spot against a light background is slightly different. Such an object can only be seen if it subtends an angle of at least 30 seconds of arc. Here the light edge surrounding the dark spot can be considered a series of light spots surrounding a dark central area. As the diameter of the dark area is made progressively smaller, the light from the edge of the point spread functions of the surrounding light point objects begins to spill over onto the central dark spot. This will eventually render the dark spot undetectable due to insufficient contrast between the central dark spot and the surrounding light region. Again, the task has become one of contrast discrimination. It

is reasonable to argue that the task of detecting a dark object against a bright background better fits the requirements of an acuity test because, unlike the reverse case, a critical size can always be found below which a dark object cannot be detected against a bright background. The fact that spatial performance for detection tasks of this type is dependant upon contrast sensitivity has been demonstrated experimentally[32]. Whether for dark objects on a light background or light objects on a dark background, detection tasks of this kind are not the ideal method of assessing spatial discrimination. This is because in the former, the task becomes one of contrast sensitivity, and in the latter, the task becomes independent of stimulus size providing the intensity is great enough.

2.5.2. *Separation (or resolution) acuity*

This is the ability to distinguish two or more objects as being separate. A simple example would be the ability to perceive two spots as being separate from one another rather than a single spot. Under optimum viewing conditions, two dark spots on a light background (or two light spots on a dark background) can be resolved as separate once they are separated by around 30 to 60 seconds of arc. According to the Rayleigh criterion, the images of two point objects can be just resolved as being separate when the centres of their point spread functions are separated by half the radius of the Airy disc (see section 2.4.1). This is illustrated diagrammatically in Fig 2.5.1.

Figure 2.5.1. The 'Rayleigh criterion' (adapted from Bennett & Rabbetts[11])



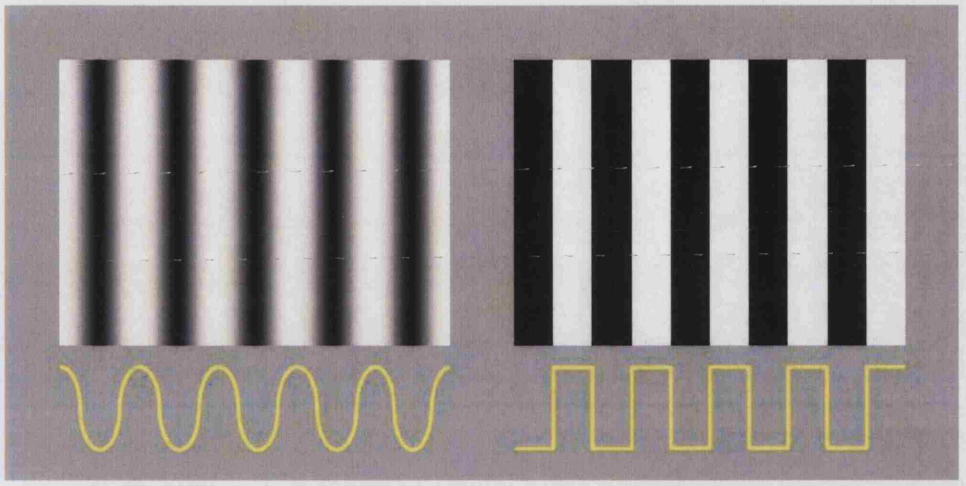
a) Retinal image distribution for two point objects which are resolvable as separate – Airy discs separated by more than half their width
b) Retinal image distribution for two point objects which are not resolvable as separate – Airy discs separated by less than half their width
c) The limit of resolution according to Rayleigh – retinal image distribution for two point objects just resolvable as separate – Airy discs separated by half their width.

The ability to perceive two dark lines on a bright background as being separate is dependant upon the width of the lines. The minimum detectable separation reduces as the lines become broader and the task approximates the detection of a single bright line on a dark background[33]. In the same way as for the case of a single spot of light on a dark background (section 2.5.1) infinitely small separations between two broad dark bands may be detected provided the background intensity is sufficiently great. The minimum separation between two light bars on a dark background is also dependent upon the width of the bars, and in the same way as for a dark spot on a light background, a minimum separation can be found below which the dark gap between the bars can no longer be detected. However, whereas the minimum detectable separation between light spots is around 30 to 60 seconds of arc, two broad light bars on a dark background may be perceived as separate when they are only around 1

second of arc apart (provided the bars are sufficiently long - about 1 degree of visual angle)[32 34].

Although the separation between pairs of objects has been used as a task to evaluate the quality of optical systems, a more common type of separation stimulus for acuity measurement is the grating. Gratings are made up of alternating light and dark bars of equal width and the luminance across the stimulus is usually modulated either sinusoidally or as a square wave function (see Fig 2.5.2). For a grating stimulus, the task is to distinguish the grating from a uniform background of the same space-averaged luminance. The difficulty of a grating is generally expressed as a spatial frequency in cycles per degree, a cycle being the distance between the centre of one dark bar to the centre of the next dark bar. One bar on a 30 cycle per degree grating subtends an angle of 1 minute of arc. A much larger angular subtense of the bars is required for the grating to be resolved than would be required for the detection of a single bar against a bright background. A potential problem with the use of gratings is the tendency of grating stimuli to underestimate the visual deficits produced by conditions such as amblyopia[35-38]. Also situations may occur with normal subjects where a spatial frequency is found at which the grating can no longer be resolved, and yet a further increase in spatial frequency allows the grating to be resolved once more. This 'spurious resolution' may occur at least in part due to aliasing effects which occur with spatial frequencies which exceed the sampling frequency of the retinal cone distribution[39]. Such spurious resolution may also occur with defocus[40]. Tests which employ more complex stimuli based upon the grating principle may however be less susceptible to this bias[41 42].

Figure 2.5.2. Sinusoidal and square-wave gratings



The yellow line describes the spatial luminance profile of the grating

In general, some of the limitations of detection acuity in capturing the essence of visual acuity (section 2.5.2) also effect separation acuity. In a clinical situation, such conceptually pure tests as two-line or grating resolution are generally deemed less practicable than optotype acuity[40].

2.5.3. *Hyperacuity*

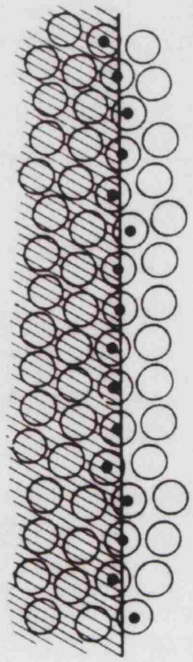
Certain spatial discriminations can be made when the threshold is much lower than normal acuity. The most common example of this form of 'hyperacuity' is vernier acuity in which an object may be recognised as misaligned relative to another when the angular displacement is only a few seconds of arc (Fig 2.5.3). Such fine degrees of discriminability would seem to be at odds with the optical and neural basis for acuity as discussed in section 2.3. This may be explained however through a refinement of Lotze's theory of local sign[43]. If we consider that each foveal cone corresponds to a given direction in space, then averaging the local signs of multiple cones on which an edge falls allows the edge to be localised to less than a cone diameter (Fig 2.5.4). This allows finer discriminability than would be predicted based upon the sampling frequency of the retinal mosaic, but requires sophisticated neural processing[40]. Vernier acuity tests are seldom employed either in clinical practice or clinical research. This may be due to the fact that the measure of vernier acuity is the variance of repeated alignment settings performed by the subject, and as such they are inherently time

consuming. It has been suggested however that such tests may have a role in predicting post-operative acuity in subjects with dense cataract preventing visualisation of the retina[44-46].

Figure 2.5.3. An example of a vernier hyperacuity task

The subject adjusts the lateral position of the lower line such that it appears to be in line with the upper line

Figure 2.5.4. Localising an edge to less than a cone diameter in a hyperacuity task

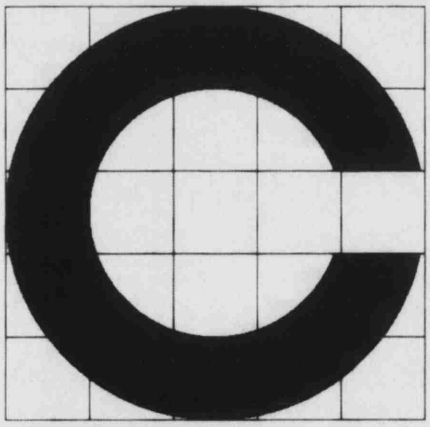


The local signs of those cones on which the edge falls are indicated by dots

2.5.4. *Recognition acuity*

Recognition tests assess the smallest size at which optotypes (letters, pictures or symbols) can be recognised[47]. Optotypes are typically based on a grid which dictates the proportions of the elements of the stimulus relative to the size of the whole stimulus. For example a Landolt-C is based upon a 5x5 grid such that the stroke width, and the width of the gap in the ring both subtend one fifth of the total size of the optotype (see Fig 2.5.5). The relative dimensions of the optotype are kept constant, but the angular subtense is altered to vary difficulty. Most visual acuity measurements feature optotypes of this kind, the most common format being that of the acuity chart which typically features rows of letters of varying sizes. Although recognition acuity is a less direct test of resolving power than separation or resolution acuity, it benefits from not requiring a binary decision which is subject to guessing. Letter optotypes belong to a moderately sized collection of well-known patterns and optimise information transfer between patient and subject[40].

Figure 2.5.5. Landolt-C optotype based upon a 5x5 grid



The acuity chart has remained the most common test of acuity in clinical practice since its introduction in 1862[48] (see section 2.10). Its dominance has been reinforced by publication of various standards relating to acuity measurement both in clinical practice[49 50] and clinical research[51]. The widespread familiarity of the general public with visual acuity charts is a distinct advantage of this type of test. Indeed, a visual acuity measurement is probably one of the few psychophysical tests which can be

happily performed by most people without any explanation. Nonetheless, it should be remembered that a subject undergoing any form of recognition acuity measurement is performing a complex cognitive task. The image of the test stimulus produced on the retina has to be resolved, interpreted, compared with a series of possible responses from previous experience, and once a likely match has been decided upon the response has to be communicated to the examiner. The complexity of this process may increase the scope for inherent measurement error.

2.5.5. *Objective measures of acuity*

The cognitive and verbal demands of performing a recognition acuity test as described in section 2.5.4 are such that they will exceed the abilities of some subjects. For this reason, several objective tests of acuity have been developed which illicit a subconscious response to an acuity stimulus:

Visually evoked potentials

This technique presents a spatial pattern such as a grating or checkerboard to the subject and uses changes in the electroencephalogram (EEG) as evidence that the pattern was resolved[52]. The spatial frequency of the stimulus is then altered to establish a threshold from which the subjects resolving power can be deduced[53]. The EEG is subject to a degree of background 'noise' which typically requires the averaging of a series of measurements to improve the signal to noise ratio. This technique allows a good assessment of resolution, although it should be remembered that the neural processing required employs a different neural pathway to that utilised during optotype acuity measurements, and as such the results may not always be comparable[54].

Involuntary eye movements

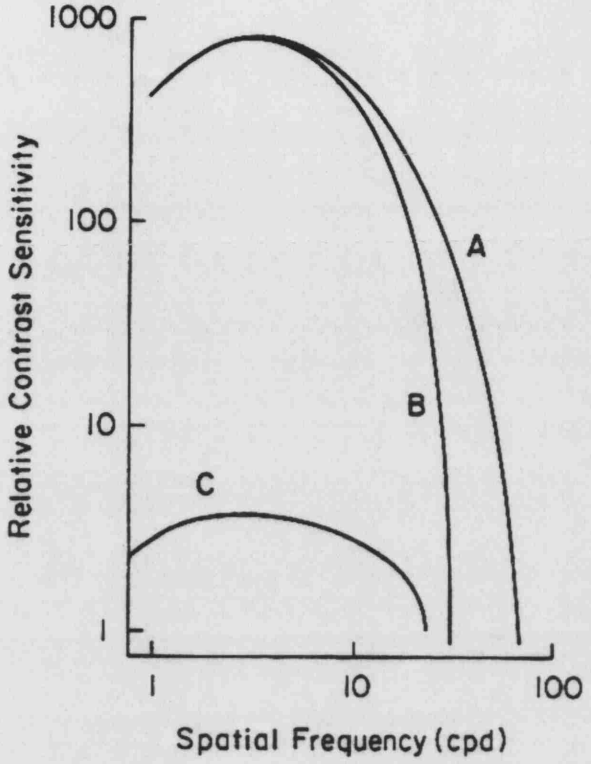
Some forms of eye movements are initiated in response to a change in afferent visual input, and have no significant voluntary component. Pursuit and optokinetic eye movements are of this type. This has been used to form the basis of acuity tests in which successful resolution of the stimulus is indicated by eye movement. For example the spatial frequency of vertical black and white stripes on a rotating drum can be varied to determine the upper limit of spatial frequency which initiates optokinetic nystagmus[55].

Objective tests of acuity form part of the armoury of tests used to assess visual acuity in children. Paediatric visual acuity tests are discussed in section 2.10.14.

2.5.6. *Contrast sensitivity*

A contrast threshold is the smallest degree of contrast which is required to distinguish a target from a uniform background. Contrast sensitivity is the reciprocal value of the contrast threshold, such that a high degree of contrast sensitivity equates to the ability to detect very faded or washed-out images against a uniform field. As referred to in section 2.2, images are often considered as complex waveforms consisting of various component spatial frequencies[4]. Contrast sensitivity in the normal visual system varies as a function of spatial frequency as shown by the normal Contrast Sensitivity Function or CSF (curve A in Fig 2.5.6). Whereas the measurement of high contrast visual acuity only gives information about the performance of the visual system for high spatial frequencies, the CSF gives a more complete description of visual function. Contrast sensitivity is maximum for spatial frequencies in the middle of the range. The slight drop off towards low spatial frequencies reflects the effects of lateral inhibition during image processing, whereas the dramatic decline in sensitivity towards the high spatial frequency end is determined predominantly by the optical characteristics of the ocular media[57]. Contrast sensitivity tests attempt to describe more of the CSF curve than visual acuity which only predicts the intersection with the X-axis. This is of particular use in those conditions which may differentially affect sensitivity at lower spatial frequencies leaving visual acuity at near normal levels e.g. cataract, amblyopia, multiple sclerosis, mild toxic neuropathies, glaucoma and ocular hypertension, Parkinsonism, intracranial hypertension, macular degeneration and some other forms of retinal disease[58-68]. This effect is illustrated by curves B and C in Fig 2.5.6. The contrast sensitivity function described by curve C is consistent with a much poorer level of visual function than that represented by curve B. However, because the intersection of these two curves with the X-axis is similar, the respective visual acuities will also be relatively similar. Therefore a measure of high contrast visual acuity will be relatively insensitive to the abnormality causing this reduction in visual function.

Figure 2.5.6. The Contrast Sensitivity Function (after Simons[56])



A – Normal contrast sensitivity function

B – High frequency attenuation

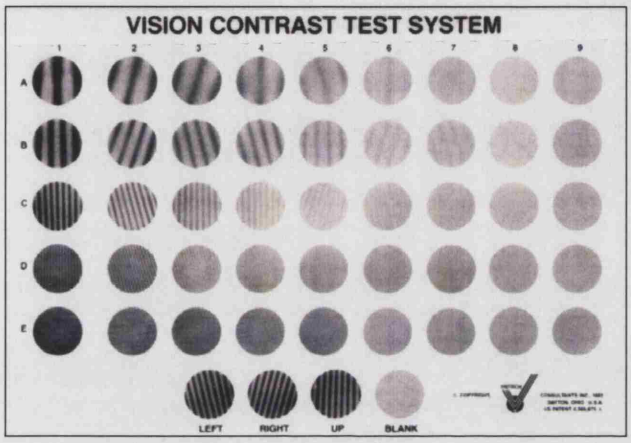
C – Diffuse reduction

A formal assessment of the full CSF is a time consuming procedure typically carried out using grating stimuli. More clinically orientated tests which do not measure the whole curve have been developed. Fig 2.5.7a shows the Vistech contrast sensitivity chart which does allow an approximation of the full CSF by estimating contrast sensitivity at several different spatial frequencies. The Pelli-Robson chart shown in Fig 2.5.7b is specifically designed to compliment visual acuity data by measuring contrast sensitivity at single relatively low spatial frequency. Contrast sensitivity has failed to live up to the high expectations created following the initial investigations of the 1960s and 1970s[69]. This may be in part due to the remarkable fact that although high contrast optotype acuity describes only a small part of the CSF, cases in which a central visual deficit is not revealed by high contrast acuity alone appear to be very much the exception rather than the rule. Contrast sensitivity does however provide valuable

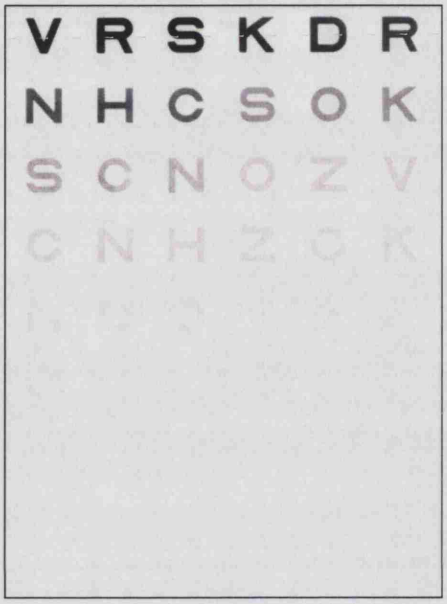
quantitative data concerning visual performance which may supplement the information gained from high contrast visual acuity measurements[69].

Figure 2.5.7. Two clinical tests of contrast sensitivity

a) The Vistech chart



b) The Pelli-Robson chart



2.6. The Importance of Visual Acuity

Visual acuity is the most commonly performed test of vision and undoubtedly the most important. For the vast majority of ophthalmic and optometric clinic visits, visual acuity will be the only measurement of visual function undergone by the patient. Despite the fact that visual acuity is only one aspect of visual function (see section 2.3), it appears to provide a sufficiently good overview of visual function to be the only routine test of visual function. Numerous other tests of visual function are invaluable in the diagnosis and management of eye disease, but all are applied selectively according to the presence or absence of additional symptoms, risk factors and/or clinical findings. In view of its predominance in clinical practice, it is not surprising that its relevance extends beyond clinical practice and clinical research.

2.6.1. *Detecting abnormality*

The measurement of visual acuity is fundamental to any routine examination within ophthalmic and optometric clinical practice. The measurement of visual acuity is an essential step towards confirming the integrity of the visual system. If good acuity can be demonstrated in both eyes, the likelihood of an abnormality affecting the visual system is reduced considerably. Combining this information with a simple ophthalmoscopic examination rules out the vast majority of abnormalities which may affect the visual system. In addition to its role in detecting abnormality in a clinical environment, visual acuity has been a core element of vision screening programmes such as those employed within schools[70].

2.6.2. *Statutory*

The predominance of visual acuity measurements in the clinic is replicated outside the clinical environment where visual function is described for statutory or other legal purposes. For example, eligibility for blind or partial sight registration is typically judged according to the level of visual acuity[71]. In some cases a standard may combine visual acuity with an assessment of the visual field to provide a more comprehensive description of visual function, e.g. the requirement for a UK driving licence[72]. In this context, visual acuity is used to help determine whether someone's visual function is

sufficient to allow them to function either for activities of daily living in general, or for a specific task.

2.6.3. *Occupational*

Various occupations have a visual standard which successful job applicants are required to meet. Here, as for the previous category, visual acuity is being assessed to predict whether a person possesses sufficient resolving power to be able to function in a certain context, rather than establishing whether or not their visual system is normal. Examples of occupations for which there is a visual requirement include the armed forces, electrical engineering, civil aviation, prison officers, and the coastguard. Again visual acuity may be combined with other tests of visual function as in the colour vision requirement for entrance into the Royal Navy.

2.6.4. *Monitoring disease*

Visual acuity is often used to monitor vision over time. This may be to establish the progression or regression of disease of the visual system, or to evaluate the effect of treatment. The important issue of detecting change is discussed in more detail in section 2.9.

2.6.5. *Surveys*

Population-based surveys of eye disease will invariably utilise a measure of visual acuity. This may either be to screen for refractive error or eye disease, or to establish the prevalence of visual disability in a certain population. In many cases, a pass/fail criterion may be used to decide which individuals merit a more comprehensive examination.

2.6.6. *Refraction*

Visual acuity is acutely sensitive to uncorrected refractive error through its selective filtering of high spatial frequencies. Accordingly, visual acuity charts are invariably used during the process of refraction to aid in endpoint determination.

2.6.7. *Research*

The role of visual acuity in clinical research is varied and may overlap with several of the categories referred to above. Examples of potential uses of visual acuity in research include:

- To establish eligibility for entry into a study
- To establish the effect of a disease upon visual function
- To establish the course of a disease over time
- To establish the strength of a treatment effect
- To establish the visual side effects of a treatment
- To establish the impact of visual impairment on a population
- As a potential risk factor

The aspects of visual acuity test performance which are most critical will depend on the use to which they are to be put. Section 2.7 describes the various aspects of acuity test performance. Section 2.8 discusses which aspects of test performance are pertinent to the various potential applications referred to in this section.

2.7. Desirable Characteristics of Visual Acuity Tests

This section will attempt to outline those characteristics which would be desirable in the perfect acuity test.

- **Accuracy**: This is the degree to which an estimate represents (on average) the true value of what is being measured[73]. Inaccuracy therefore indicates the presence of bias leading to a consistent over- or under-estimate of acuity (see section 4.1.1 for more detail).
- **Precision**: This represents the repeatability of a test. In statistical terms, it represents the inverse of the variance of a measurement or estimate[73] (see section 4.1.2 for more detail).
- **Precision should be independent of acuity**: If the degree of precision does not vary with acuity, then simple numerical cut-offs (e.g. a criterion for significant change) can be used regardless of the underlying level of acuity.
- **Visual acuity data should be easy to analyse**: This facilitates audit as well simplifying clinical research.
- **Universally applicable**: The test should be suitable for use with all levels of vision & all capabilities (including children & those with learning difficulties or a different language).
- **Quick to administer**: The test should also be fast to perform and easily understood by examiner and subject.
- **Inexpensive & robust**: The test should be cheap enough such that it is practical for use in all of the situations outlined in section 2.6. It should also be sufficiently robust to be able to function normally in either a clinical environment, or in the field.

2.8. Characteristics Pertinent to the Various Uses of Visual Acuity Tests

The most important aspect(s) of a visual acuity charts performance will depend upon the reason for taking the measurement. Having considered the possible applications of acuity measurements, and the various aspects of test performance, we can now consider which characteristics are pertinent to each potential application (Table 2.8.1).

Table 2.8.1. Characteristics pertinent to the some of the uses of visual acuity tests

Potential use	Accuracy	Precision	Speed
Detecting abnormality	+++	++	++
Social/occupational	+++	++	+
Monitoring disease	+	+++	++
Surveys	+++	+	+++

Two key characteristics whose relative importance will vary according to the purpose of the test are accuracy and precision. Accuracy has been defined in section 2.7 as one which *on average* produces an estimate which is close to the true visual acuity. Accuracy is important for any situation where acuity measurements are being used to categorise people. In this situation, an inaccurate test will produce an excess of either false-positives or false-negatives depending on the direction of bias. Accuracy is therefore desirable for detecting abnormality, and for vision surveys. It is less important, however, for monitoring either disease progression or the efficacy of treatment, as the degree of bias will be present in both the base-line measurement as well as any subsequent measures. It will therefore not affect the evaluation of change over time.

The precision of a test is also important in some situations where it is necessary to categorise people. Let us consider the case of a highly accurate test (which on average will produce an estimate of acuity which is equal to the true underlying acuity). The importance of precision is most obvious if we consider an acuity measurement taken on a given individual to be a sample of size one taken from a theoretical distribution of

infinite repeated measures of visual acuity on that subject. The less precise the test, the wider this distribution, and the greater the chance that a single acuity measurement will depart from the true underlying acuity (equal to the mean of the distribution). Hence when categorising people based on a single estimate of acuity, despite having a perfectly accurate test, an imprecise test will result in a number of false positives and false negatives (the numbers of which will be equal because the test is accurate).

Whether precision is important with respect to categorising people depends on the specific reason for the categorisation. For example, if a single measure of acuity is used to determine which side of the Snellen 3/60 cut-off applicants for blind registration fall, the more imprecise the test, the larger the proportion of inappropriate categorisations. Hence in this instance, precision is important. However, if a single measure of acuity is being used within an epidemiological survey to estimate the prevalence of visual impairment in a certain population, then providing the test is accurate, precision is relatively unimportant. The reason for this is that imprecision will increase the number of misclassifications equally in both directions (false positives and false negatives). Therefore, because a large number of individuals is being screened, the estimated proportion of individuals falling above and below the cut-off will be unaffected. This is one of the few applications of visual acuity tests for which precision is relatively unimportant.

A precise test is very important where an acuity test is used to monitor disease progression or the efficacy of treatment. The greater the degree of imprecision, the larger a clinical change must be before the acuity test can be certain of detecting it. This issue is discussed further in section 2.9. The precision of an acuity test may be improved by using a finer scale increment (see section 2.10.6). However, it should be noted that this may create a trade-off between the precision and duration of an acuity measurement require more stimuli to be attempted which in turn will increase the duration of the test. As well as the obvious implications for clinic resources, this may also have implications for subject fatigue.

The requirements of an acuity test which is being used for the purposes of refraction only are minimal. Neither accuracy nor precision are required. Optical defocus selectively filters out high spatial frequencies thereby removing them from the retinal image. As high spatial frequencies are only present in sharp edges, uncorrected

refractive error will blur the edges of any target, regardless of size. A good observer undergoing refraction requires little more than a single high contrast target with sharp edges, to differentiate between the degree of blur produced by two different spectacle lenses. Observers who are less acute in their subjective responses often require a target which is close to their threshold as well as possessing sharp edges. This gives the cue of loss of legibility, in addition to loss of clarity when optical defocus occurs. Hence, the requirements of a visual acuity chart for the purposes of refraction alone are limited to a selection of high contrast, sharp-edged targets. It is desirable to have targets of varying sizes such that most subjects will have a near-threshold target which will be rendered unrecognisable by a small degree of blur. A selection of targets at each size is preferable such that a reported loss of legibility can be checked using a stimulus which the subject has not already seen.

2.9. DETECTING CHANGE

The primary use of a visual acuity test is probably to distinguish between normality and abnormality. Another very important use of these tests is to detect change in a subject's clinical status. As mentioned above, the precision of a visual acuity test influences the ability of a test to detect change. This section will consider the importance of precision both to detection of change in individuals and the detection of change in groups.

2.9.1. *Detecting change in groups*

Many investigations in clinical research involve determining the influence of a condition or intervention upon a subject's clinical status. The variability inherent in biological systems often prevents unequivocal conclusions being drawn from observations carried out on a single subject. It is therefore commonplace for a number of appropriate subjects to be observed and the results interpreted collectively. This reduces the impact of random variability upon the results. Clinical tests are subject to such random error often referred to as test-retest variability (TRV). Precision (see section 2.7) is the inverse of TRV, such that a precise test is associated with very little TRV. Where a clinical test is being used to monitor a group of subjects, the presence of TRV will tend to obscure true changes in clinical status such that the detection of a given degree of change will be delayed. Also if the test is used to compare one or more groups of subjects, TRV will tend to obscure genuine differences between the groups such that only larger differences will be detected. The detrimental effect of TRV on detecting change or detecting differences between groups may be offset by increasing the number of subjects being studied. This, however has implications for the cost of clinical research.

Because of the uncertainty introduced by random variability, statistical methods are required to determine the probability e.g. that an observed difference between groups is due to chance. If the probability of the observed difference occurring by chance alone is sufficiently small then the measured difference can be attributed to a true change in clinical status. A detailed discussion of the statistical methods used when

comparing groups is beyond the scope of this section. However, section 4.1.6 of this thesis describes the 't-test' which is commonly used for this purpose.

2.9.2. *Detecting change in individuals*

The detection of change in individuals is a very common task for a visual acuity test in routine clinical practice as well as being used in clinical research. In the same way as for detecting differences between groups, the detection of change in individuals is compromised by the presence of TRV. The less precise the test, larger the degree of TRV, and the larger measured change must be before it can be attributed to a true change in clinical status.

Whereas the statistical methods for seeking differences between groups of subjects are quite well established, less of a consensus exists on what method should be used to detect change in individuals. It is likely that in clinical practice clinicians use their experience to judge whether a measured change in acuity is sufficiently large to be due to a true change in clinical status. In fact it appears that the use of clinical experience to judge what size change should be regarded as genuine has sometimes extend into the field of clinical research. For example, whilst reviewing the use of visual acuity measurements in trials of anti-cataract formulations, Elliott[74] highlighted three published studies (those of Ichikawa et al[75], Libutti and Pannarale[76], and Sanders et al[77]) all of which considered a 2-line change in acuity to be significant, without providing any rationale behind this choice of cut-off.

There are a number of statistical techniques which might be used to assess the significance of measured change (e.g. linear regression analysis). These however require the use of a computer which may be impractical for routine clinical use. A more practical alternative is to use an estimate of precision in the form of the 95% test-retest range (TRR). This is the range within which 95% of measured differences between test and retest will lie in the absence of a true change in clinical status (see section 4.1.2 for more detail). The width of the 95% TRR is used as a cut-off such that measured change which exceeds the 95% TRR is deemed to be the result of a true clinical change, whereas measured change equal to or less than the 95% TRR is

attributed to TRV. This method has been widely advocated for use with visual acuity scores[74 78-81] as well as for contrast sensitivity[82-84], and colour vision[85].

In summary, the higher the test-retest variability or TRV of visual acuity measurements (i.e. the less precise the test) the more the detection of change will be compromised. This may lead to unnecessarily large sample sizes in clinical trials, and may delay the recognition of disease progression (or treatment effect) in individual subjects. Statistical methods may be used to deal with the uncertainty encountered when attempting to determine whether measured change is due to a true change in clinical status, or TRV alone.

2.10. The Development of Optotype Acuity Measurement

2.10.1. *Overview*

The assessment of visual function in terms of resolving power can be traced back thousands of years. The Egyptians have been credited with the earliest eye test which had to be passed by anyone wishing to go hunting[86]. They tested visual function through the ability to perceive as separate two stars within the constellation we now refer to as 'The Plough'. A precedent of primarily using reading tests to assess visual function dates from before the Middle Ages culminating in the set of reading samples produced by von Jaeger in 1854 for the documentation of functional vision[87]. These samples proved popular and are still widely used in North America. They lack a scientific form of notation as the Jaeger numbers are derived from item numbers in the Viennese printing house catalogue where they were first produced. Also, the lack of a standard font led to numerous different versions of the samples, as those wishing to use the test attempted to replicate the original using print samples from their local printer[88]. The first attempt at a transition to charts viewed from a distance was made by Kuechler, a German ophthalmologist who published a series of such charts in 1843[89]. However, this approach did not become popular until advocated by Donders, for whom distance targets for visual assessment were an essential part of his search for a more scientific approach to refraction[2]. He coined the term 'visual acuity' to describe the "sharpness of vision" and charged an ophthalmologist colleague, Herman Snellen, with developing the first chart for it's measurement. Snellen's chart (which he referred to as his 'Optotypes') was published in 1862[90]. Snellen considered various types of stimuli (Fig 2.10.1) before deciding on letters and numbers (Fig 2.10.2). In keeping with Donders' ideas on visual acuity, Snellen notation was defined as the ratio between measured visual performance and a standard level of performance (see section 2.9.5). Snellen defined 'standard vision' to be the ability to recognise a letter target when is subtends 5 minutes of arc at the eye (the letter target being based upon a 5x5 grid such that the standard letter features a stroke width which subtends of 1 minute of arc). This choice appears to have been based upon previous work (by the English astronomer Robert Hooke, and physicist Herman von Helmholtz, among others) which

showed the minimum separation at which the human eye can resolve two objects as being separate to be 1 minute of arc.

Figure 2.10.1. Two of Snellen's experimental optotypes

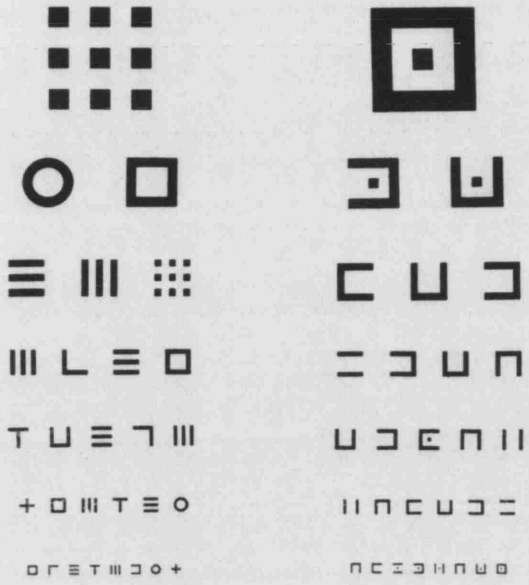
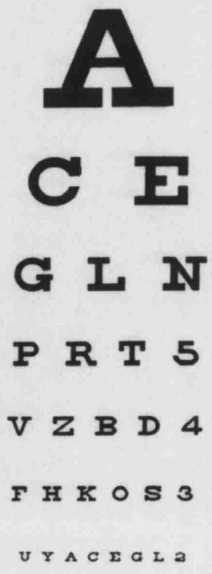
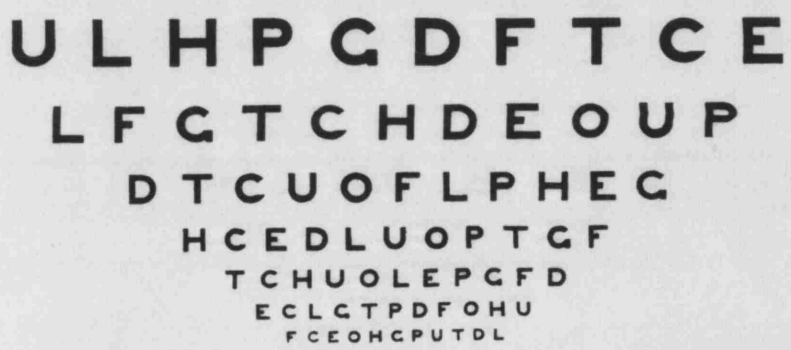


Figure 2.10.2. Snellen's chart as published in 1862



Snellen's chart was first published in 1862[48], followed by an English version in 1864[91], and is still widely used today. Recent decades have seen much criticism of Snellen's chart design[74 92-94] (see section 2.12). In view of this, it is interesting to note that some of the earliest suggestions for improving on Snellen's design were made as early as 1868 by American ophthalmologist John Green[95]. Green incorporated his suggestions into a chart (Fig 2.10.3) which subsequently failed to gain acceptance within the ophthalmic community. The merits of several of Green's proposals (including the use of san-serif letters, similarly legible letters, proportional spacing between letter stimuli, a geometric progression of letter sizes, and a non-truncated design) are now recognised as essential parts of a scientific approach to visual acuity measurement[51].

Figure 2.10.3. A section of Green's chart

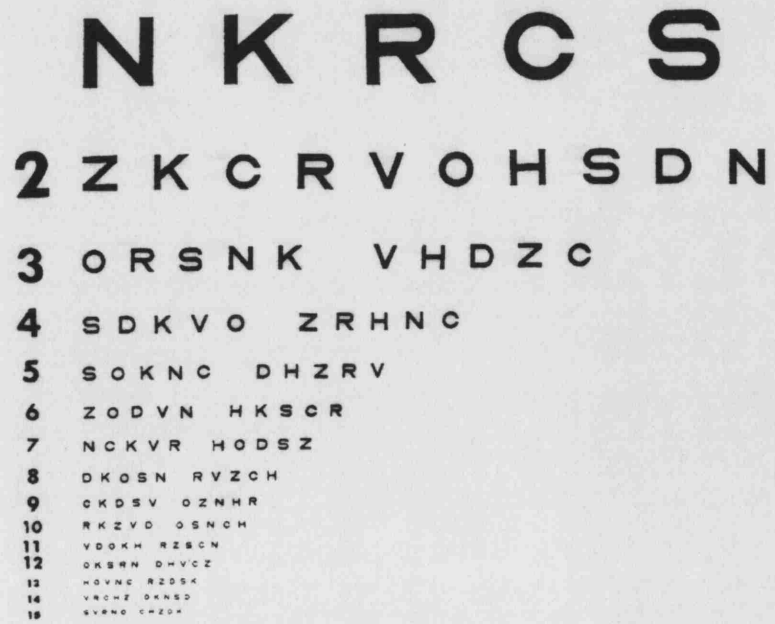


The continuing predominance of the Snellen chart in clinical practice is evidence of the fact that the twentieth century was slow to embrace any developments in the assessment of visual acuity. This lack of progress was probably due, at least in-part, to a lack of recognition of the limitations of the Snellen design. The main requirements for a clinician during the first part of the twentieth century, would have been a chart which would act as a target for subjective refraction (for which the requirements are minimal - see section 2.8), and which would allow abnormal levels of vision to be differentiated from normal levels of vision. Snellen's chart was sufficient for both these functions, and as such there was little desire for an alternative method. One notable development was that proposed by Landolt in 1889 (cited by Cowan[96]) who, realising that Snellen's optotypes were not equally recognisable, proposed the use of a

single optotype whose orientation is varied. Although acceptance of the 'Landolt-C' was limited in clinical practice, it has long been employed as a standard in research, and has been incorporated into several standards concerning visual acuity measurement[50] [97]. It has been suggested that the preference for letter optotypes in clinical practice is because testing with letters is simpler and quicker[98].

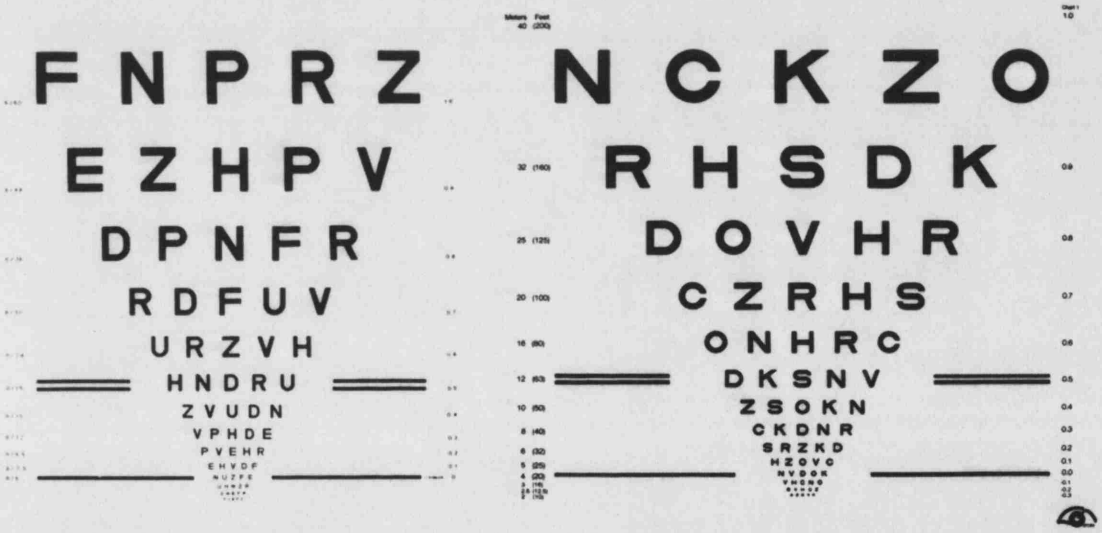
Interest in the measurement of visual acuity was reinvigorated in the second half of the twentieth century with increased interest in the rehabilitation of those with visual impairment. A scientific approach to visual rehabilitation merited a move away from a dichotomous classification of 'sighted' and 'blind' towards a more graduated classification of visual disability. This need was reinforced by the World Health Organisation who revised the International Classification of Diseases[99] following a survey of national blindness definitions[100]. In 1959 Sloan developed a series of 10 optotypes whose legibility is approximately equal, as well as being equal on average to that of the Landolt-C[101]. She produced a chart which featured all 10 letters on each line, along with a geometric size progression, and both Snellen and angular letter size notations.

Figure 2.10.4. A version of Sloan's 1959 chart



A seminal contribution to the design of visual acuity charts was made in 1976 when Bailey and Lovie published a new chart featuring a novel design in which each line contained five letters, and the spacing between letters and lines was proportional to the letter size on that line[92] (Fig 2.10.5). They used another set of approximately equally legible letters published earlier by the British Standards Institution[49]. The novel chart layout was designed to produce a consistent degree of contour interaction across the chart. The Sloan letter set and the Bailey-Lovie design were united in 1982 when Ferris and colleagues incorporated both into a letter chart for the standardised measurement of visual acuity in the Early Treatment Diabetic Retinopathy Study (ETDRS)[94]. The Bailey-Lovie/ETDRS design has since been advocated by various standards internationally[51 102].

Figure 2.10.5. The Bailey-Lovie and ETDRS charts



2.10.2. *Choice of letter*

Snellen's original chart used a type face known as 'Egyptian Paragon' and featured pronounced perpendicular bars or 'serifs' at the termination of the letter strokes. He realised soon after his chart was published that the letters used were not all equally recognisable. Attempts to minimise letter legibility (or difficulty) as an extraneous variable began as early as 1868 when Green proposed an alternative set of letters to

those used by Snellen suggesting that they were 'simpler in construction' and 'more nearly equal'[95]. Unfortunately, the lack of acceptance for his proposals was probably due, at least in part, to aesthetic considerations, as his letters lacked serifs and were criticised for looking 'unfinished'[47]. Dennet[103] used a different approach in making small adjustments to the size of letters according to their legibility, to produce more uniform difficulty. Some authors suggested mixing letter styles to achieve more uniformity[96 104]. Again the resultant lack of aesthetic appeal was likely to be the main reason for their lack of success[47]. Banister[105] used an interesting approach later repeated by Finkel[106] whereby only those letters which are easily confused with one another are employed. These are further subdivided into easy or difficult groups. Each line on Banister's chart includes a small number of difficult letters. The subject is only allowed to progress to the next line if they correctly name both the easy and difficult letters. Should the easy letters be correctly named but not the difficult letters, the measurement is terminated at that point. A different approach was advocated by McMonnies whereby partial credit is given if the subject responds with a letter which is similar in appearance to the stimulus (e.g. P for F and Y for V)[107]. Various other workers carried out investigation into relative letter legibility including Coates[108] and Woodruff[109]. However, the two letter subsets employed in the current gold standard charts are those of the British Standards Institution[49] in the case of the Bailey-Lovie chart[92], and Sloan's set of letters[101] in the case of the ETDRS chart[94]. Even this seemingly rigorous approach has received criticism e.g. from McMonnies who has suggested that using different sequences of the same 10 letters on each line will improve precision still further[110]. Using the British Standard letter set, McMonnies has also demonstrated that the ranked order of legibility within a letter set may vary according to the test within which they are used[111].

2.10.3. *Alternative 'Optotypes'*

Snellen's original chart was referred to as Snellen's 'Optotypes', a word which has since become the collective noun for the stimuli used on visual acuity charts: usually letters or symbols. The main alternative to letters are optotypes such as the 'Landolt-C' (cited by Cowan[96]) and Snellen's 'Illiterate-E'[112] (now referred to as the 'Tumbling-E'), whose orientation is elicited rather than the name. These optotypes benefit from an

inherent similarity between the various stimuli used. The Landolt-C was first specified as a standard in 1909[97] a recommendation which has been endorsed more recently[50]. However despite being suggested as the gold standard optotype against which all others should be judged, it is perhaps seldom used in modern clinical practice except where the subject is unfamiliar with the Roman alphabet[A].

Sloan[113] raised several objections to the use of the Landolt-C as an acuity stimulus. She felt that it tended to overestimate acuity in the presence of astigmatic optical defocus. She also felt from personal experience that an acuity measurement using a Landolt-C chart was considerably more demanding of the examiner's attention than one using a letter chart. In defence of this opinion she quoted the American Medical Association who's committee has recommended that the Landolt-C be reserved for research purposes, whilst advocating non-serif letters whose difficulty has been calibrated against that of the Landolt-C for routine clinical use. A similar opinion concerning the Landolt-C was voiced by America's National Academy of Sciences-National Research Council (NAS-NRC) Committee on Vision[50]. They also adopted the Landolt-C as a standard whilst recommending the continued use of letter stimuli for visual acuity charts due to the difficulty in tracking responses to Landolt-Cs, and the increased scope for ambiguity in the communication of subject responses. Another potential drawback of using single optotypes whose orientation is varied (as opposed to multiple optotypes) relates to the reduced number of possible responses. Contemporary approaches to scoring visual acuity charts involve the use of forced choice paradigms in which the subject is required to attempt the letter whether or not they believe they are capable of a correct response (see section 2.10.5). Where only 4 possible responses exist, as in the case of the 'tumbling-E', the chance of guessing the correct response are much greater than if letter stimuli are used (where 26 alternatives are available). The consequences of this increased chance of a correct guess have only recently received detailed consideration[114]. This issue is less of a concern for the Landolt-C if it is used in one of 8 orientations, rather than 4.

[A] Landolt Cs are used in routine clinical practice in Japan due to the fact that the characters of the Kanji alphabet are too complex for clinical testing.

The practical difficulties associated with the use of tumbling-E and Landolt-C charts may have contributed to the recent proliferation of visual acuity charts catering for non-English speakers from various parts of the World[115-120]. This suggests that clinicians and/or researchers perceive a benefit in using stimuli which are familiar to the subject undergoing the measurement, as well as being amenable to an easy verbal response. In fact the benefits of familiarity have extended to the use of familiar symbols rather than the standard E and C for populations who do not utilise a formal alphabet[120].

One further factor relating to the choice of optotype is specific to paediatric vision testing. It is known that young children are subject to confusing some optotypes for their mirror image[121-123]. For this reason, some tests of children's vision use optotypes which are symmetrical about a vertical axis to remove any scope for these reversals affecting the results (e.g. Lea symbols and HOTV letter optotypes).

2.10.4. *The progression of letter sizes*

The progression of letter sizes employed by Snellen in his original chart was both arbitrary and irregular. However, it has long been accepted that the incremental change of letter size between adjacent lines on the chart should be constant. Arithmetic progressions have been suggested (such as those proposed in the 19th century by Monoyer and Oliver – cited by Bennett[47]), but more popular has been the idea of a geometric progression of letter size[47 95 97], something which Snellen himself eventually came to favour[124]. Sloan[125] proposed that resolution thresholds are similar to other visual thresholds in that they probably follow a Weber's-law like function[A]. If this is the case, then a geometric progression also is more appropriate when trying to maintain a constant degree of discriminability across the range of acuities encountered. One such scale was introduced by Snell and Sterling in 1925 when publishing their concept of Visual Efficiency[127]. This percentage scale was based on experimental evidence, and was adopted by the American Medical Association in 1955[128], and later expanded to cover other types of ocular abnormality[129].

[A] Weber's law[126] dates from 1834 and states that geometric increments in stimulus give rise to linear increments in sensation.

Most popular of the geometric scales however has been the logarithmic scale[92 94 98]. LogMAR represents the base-10 logarithm of the Minimum Angle of Resolution (MAR) and was favoured over MAR, Visual-Efficiency and the Snellen fraction by Westheimer in his search for the most appropriate means of achieving an equal-discriminatory scale[130]. Westheimer showed that 'just notable differences' were about equal across the range of acuities if a logMAR scale is used. This provides evidence that Sloan's assumption that visual resolution thresholds do follow a Weber's law like function, was a valid one. LogMAR notation has been advocated by several standards organisations[50 131] and may have been more readily embraced in clinical practice in the UK had the British Standards Institution not refrained from recommending a geometric progression of letter sizes because rejecting Snellen's empirical scale was 'too heavy a price to pay'[49].

2.10.5. *Notation*

Closely linked to the issue of letter size progression is that of notation. Snellen's notation was based upon normal vision representing the ability to resolve an optotype which subtends 5 minutes of arc at the eye. It has been suggested[86 132 133] that this precept was derived from earlier work by scientists such as Helmholtz[134] and English astronomer Robert Hooke (cited by Collenbrander[133]). There is, however, debate as to whether this standard was deliberately chosen as a screening standard which most normals should achieve[133], or whether Snellen attempted to equate his standard to perfect vision but misinterpreted this earlier work such that his standard letter was twice the size it should have been[132]. The notation was specified in the form of a fraction; where the numerator denoted the testing distance, and the denominator signified the distance at which letters of that size would subtend 5 minute of arc at the eye. Accordingly a 6/9 letter would be larger than the standard 6/6 letter, and a 6/5 letter slightly smaller. The fraction may be used with any form of unit for measuring distance. As test distances are generally specified in feet in the United States, the Snellen fraction will be expressed in feet such that 6/6 becomes 20/20, and 6/9 becomes 20/30. One advantage of the Snellen fraction is that it gives some indication as to the test conditions, but this in itself becomes a disadvantage when there is a need

to compare acuities measured at different distances. For example how does 5/7.5 compare with 6/9? This difficulty may be overcome by converting the fraction into decimal form e.g. both 5/7.5 and 6/9 become 0.67 in decimal form. The decimal form of the Snellen notation is popular throughout much of mainland Europe. The key disadvantage of the Snellen fraction is that even in decimal form, it represents ordinal rather than interval data (see Table 2.10.1), and as such is not well suited to statistical analysis[135]. The most significant advantage of Snellen notation is its universal familiarity. The significance of 6/6 is recognised by all clinicians whether or not they are in a eye-related specialty. Even amongst the general public, it is commonplace to encounter subjects who are aware that 'twenty-twenty' represents a good level of vision.

Table 2.10.1. Properties associated with the four levels of measurement.

Level	Property	Example
Nominal	Classification	Eye colour
Ordinal	Order, ranking	Pain
Interval	Ranking Equality between pairs of successive data points Arithmetic operations possible Arbitrary zero point	Temperature in Celsius
Ratio	As for interval except true zero point employed	Height

An alternative form of notation is to use the visual angle in minutes, a form of notation favoured by Sloan on account of it being 'the simplest and most direct method'[113]. The convention is to use the angular size of the smallest detail within the optotype, e.g. a horizontal bar of a letter 'E'. This form of notation is most commonly referred to as the minimum angle of resolution (MAR). Preferable to using the simple MAR is to use the logarithm to the base 10 of the MAR. This converts a geometric sequence of letter sizes to a linear scale, such that a change of one unit can be regarded as equivalent at any point on the scale. It also facilitates the statistical analysis of acuity data. Bailey coined the term 'logMAR' for this form of notation[92]. One criticism of both the MAR and logMAR notations is that better acuities are associated with lower scores.

Accordingly it has been suggested that they should be thought of as measures of visual loss rather than of vision[133]. In response to this criticism, an alternative notation has been suggested by both Bailey[136] and Colenbrander[137]. The Visual Acuity Rating (VAR)[136] or Visual Acuity Score (VAS)[137] is calculated as follows:

$$VAR = 100 - 50 \text{ logMAR}$$

Accordingly, 0.00 logMAR (Snellen 6/6) becomes 100 VAR, and +1.00 logMAR (Snellen 6/60) becomes 50 VAR. Acuities better than 0.00 logMAR (Snellen 6/6) therefore have a value in excess of 100.

Another alternative to the Snellen fraction was proposed by Snell and Sterling[127] for the purpose of quantifying visual loss for legal purposes. The visual efficiency (VE) scale arbitrarily allocates a value of 1.0 or 100% to equate with Snellen 6/6, and 0.2 or 20% to Snellen 6/60. The scale was established based upon empirical data and demonstrates the following relationship with MAR:

$$VE = 0.2^{(MAR - 1)/9}$$

The VE scale has been adopted by the American Medical Association[128] and developed such that it incorporates acuity data for both eyes, as well as allowing quantification of visual field loss and restrictions of ocular motility[129].

Despite the potential drawback of better acuities being associated with lower scores, the logMAR notation has become the main alternative to the two forms of Snellen notation, and features prominently in contemporary research publications. The preference for logMAR is reinforced by the fact that the current gold standard method of measuring visual acuity in clinical research is the Bailey-Lovie chart[92] and its ETDRS derivative[94], both of which employ logMAR notation. The benefits of logMAR notation when combined with the chart design principles of Bailey and Lovie are summarised in Table 2.10.2.

Table 2.10.2. The benefits of logMAR notation

Combines a geometric size progression and a linear scale
Well suited to analysis using parametric statistical methods
Promotes an equal discriminability scale
Facilitates manipulation of the testing distance to measure poor acuities
Allows the comparison of acuities measured at different distances

2.10.6. *The scaling and scoring of visual acuity charts*

The size of the scale increment of a measurement tool is the size of the interval between adjacent points of the scale. As mentioned in section 2.10.4, the progression of letter sizes on Snellen's chart is irregular. As the scale of the chart is dictated by the sizes of the letters on adjacent lines, the scale is also irregular. This creates difficulties when trying to specify a degree of change, as a change of two lines near the bottom of the chart does not equate to a change of two lines near the top. As such, to speak generally of changes in terms of a number of Snellen lines is less useful than would be imagined from the frequency with which it is done. For charts which feature a regular progression of letter sizes, it becomes meaningful to speak of change in terms of a certain number of lines. This is simplified further still when combining a logarithmic progression of letter sizes and logMAR notation as changes of e.g. 0.2 logMAR can be considered equivalent at any point on the chart (see also sections 2.10.4-5).

Another drawback of the measurement scale on the Snellen chart relates to the precision of the resultant acuity score. The Snellen chart is scored using a 'line-assignment' method. Using this method, the only possible acuity scores are those which precisely correspond to the stimulus sizes on the chart. Therefore, the large size intervals between adjacent lines result in a coarse measurement scale. Contemporary recommendations for acuity measurement advocate an equal number of letters per line to prevent the number of letters per line acting as a confounding variable[50 51]. Kitchin and Bailey[138] pointed out that the combination of an equal number of letters per line and a logarithmic progression of letter sizes facilitates interpolation between the acuity scores represented by the stimulus sizes on adjacent lines. Dividing the inter-line size interval by the number of letters per line gives the logMAR value of each letter on the

chart, which in turn allows acuity to be scored by the letter rather than by the line. For a chart with an inter-line size interval of 0.1 logMAR and 5 letters per line (as is the case for charts based upon the Bailey-Lovie design), each letter correctly named improves the acuity score by 0.02 logMAR. It has since been demonstrated that the finer scale associated with interpolated scoring allows a considerable improvement in precision compared with line-assignment scoring[78 80 139]. Raasch et al have suggested that making the measurement scale X times finer will reduce the standard deviation of the difference between test and retest by a factor of $1/\sqrt{X}$[139].

Contemporary guidelines for the measurement of visual acuity recommend that there should be 10 letters at every size level on an acuity chart, and an inter-line size increment of 0.1 logMAR[50]. This translates to a scale increment of 0.01 logMAR. However, this recommendation is somewhat arbitrary and appears not to have taken account of interpolation[139]. The current gold standard charts (the Bailey-Lovie and ETDRS charts) have an inter-line interval of 0.1 logMAR, but 5 letters per line resulting in a scale increment of 0.02 logMAR. Again this specification was an arbitrary one. Although several authors have shown that the precision of these charts could be improved further by reducing the size of scale increment[79 139], the upper limit of precision which may be achieved by making the scale increasingly finer has not been established.

An alternative method of scoring visual acuity measurements is by using the subjects responses to plot the probability of a correct response against visual acuity. The resultant sigmoid-shaped psychometric function can be used to determine the level of acuity corresponding to the 50% probability of a correct response[140]. This method of threshold determination is relatively time consuming and has been shown to offer no advantage in terms of precision over the interpolated scoring method[78 141].

2.10.7. *Termination rules and 'forced choice' procedures*

The traditional method of conducting a Snellen visual acuity measurement requires the subject to continue attempting increasingly smaller letters until they feel they are no longer capable of a correct response. The approach of allowing the subject to choose their own threshold criterion may bias the resultant acuity score for reasons relating

either to the subject or the examiner. One source of bias relating to the subject relates to the fact that subjects with less conservative criteria will tend to produce better acuity scores than those with more conservative criteria. Likewise, the subject who has gained confidence between their first and second attempt at the test may show an improvement merely because their increased confidence is associated with an increased willingness to attempt difficult near-threshold stimuli. The examiner conducting the acuity measurement may also be a source of bias. It is common practice for a clinician to terminate an acuity measurement when the patient has read e.g. the 6/6 line on a Snellen chart. This probably occurs because the clinician is satisfied that the level of vision is within normal limits, and yet a young patient may well have managed another 2 or 3 lines of letters[142]. Conversely, another subject struggling to read beyond the 6/9 line may receive encouragement from the clinician in the hope that the patient will achieve 6/6. In this way, a score of 6/6 may be recorded for both patients, despite the fact that their true resolving power differs considerably. These potential sources of bias have received little attention even in the scientific literature until recently[114]. It seems likely that they will have a widespread influence on the measurement of acuity in clinical practice. Interestingly, as well as compromising the usefulness of visual acuity measurements, the sources of bias described here may instil in the clinician an over-optimistic impression of the chart's performance.

A more rigorous approach to terminating an acuity measurement is to stop once fewer than half of the letters on a given line are named correctly. This approach has been employed in recent studies featuring contemporary logMAR charts where the measurement has been terminated once 3 or more mistakes have been made out of the 5 letters on a line[139 143-146]. By definition, the use of such a 'termination rule' requires the subject undergoing the measurement to attempt several letters which are below their acuity threshold. As many subjects are at least a little reluctant to do this, an additional rule can be used to require to them to keep attempting stimuli until the termination criterion is met. These so called 'forced choice' procedures removes sources of bias relating to both the subject and the examiner. Carkeet[114] has recently employed a modelling approach to show that the most effective termination rule depends on how many alternatives the subject has to choose from (i.e. for a letter chart the patient has a 26 alternative forced choice, whereas for tumbling-E stimuli, there are

only 4 alternatives). For example, the optimum termination rule for scoring acuities with the Bailey-Lovie and ETDRS charts is 4 or more mistakes on a line.

It should be noted that the use of strict scoring procedures is likely to lengthen the times required for an acuity measurement.

2.10.8. *Letter and line spacing*

Optotype legibility is dependant upon the proximity of adjacent contours. In a study published in 1963, Flom showed that the legibility of a Landolt C optotype reduces as surrounding contours are moved progressively closer to it[147]. Interestingly, as the surrounding contours become tangential to the Landolt C, the legibility increases again. Such a reversal in the relationship between stimulus legibility and the separation of adjacent contour is suggestive of a neural mechanism. Flom introduced the term 'contour interaction' to describe this phenomenon[147]. The effect occurs in normals[94 148] but is known to be more pronounced in amblyopic eyes[149-151], which is also suggestive of a neural mechanism. The greater the degree of amblyopia, the greater the influence of contour interaction upon the acuity score[150]. Interestingly, there is evidence that the detrimental effects of contour interaction may be mitigated by defocus and reduced contrast[152].

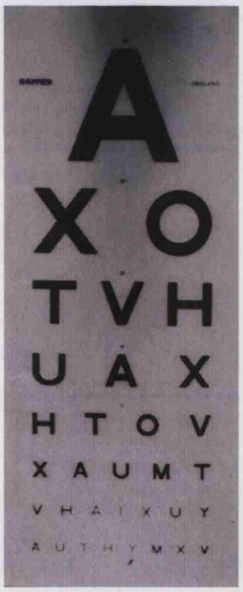
It is perhaps unsurprising in view of the design of visual acuity letter charts that they are subject to the effects of contour interaction. It is now known that both optotype[147] and vernier[153] (see sections 2.5.3 and 2.5.4) acuity measurements are subject to contour interaction. Early visual acuity charts were not standardised with respect to inter-letter spacing or the number of letters per line. This can be readily seen through inspection of a modern Snellen chart. Fig 2.10.7 shows the letter spacing on the third row to be less than one letter width, whereas the fourth row features letters which are separated by in excess of a letter width. This introduces a variable degree of contour interaction across that chart which may confound the variable of interest, letter size.

More recently, tests of visual acuity have been designed to control for the effects of contour interaction[92 94 98 154]. The spacing between letters and lines is typically specified as a proportion of the size of the letter stimuli on that line. Although the effect of varying letter spacing within visual acuity charts has had some study[139 155],

recent chart designs have letter spacing which is chosen arbitrarily[3]. This may result in acuity charts which require a larger display area than is necessary.

Stager et al[156] demonstrated that the level of contour interaction produced when letters are placed in a linear arrangement can be approximated for single letter conditions by surrounding the single letter with 'crowding bars'. Such bars have been arbitrarily incorporated into some charts to help produce consistent contour interaction[154], but their effect upon chart measurements has been subject to little investigation. It should be noted that the term 'contour interaction' is often used interchangeably with the term 'crowding'. It has been suggested, however, that crowding should be distinguished as being the resultant increase in difficulty when a closely packed arrangement of letters requires more accurate eye movements[3].

Figure 2.10.7. A version of the Snellen chart

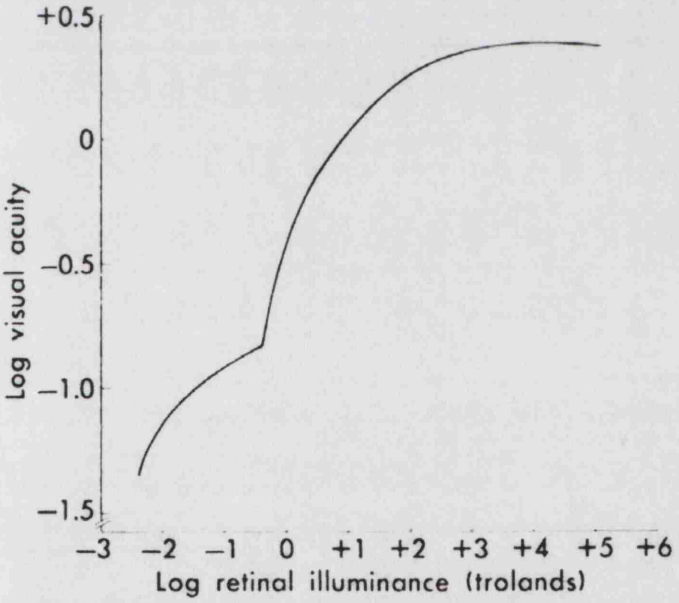


2.10.9. *Luminance and contrast*

Visual acuity increases as a function of background luminance from mesopic to high photopic levels[157]. However, although the increase in visual acuity is considerable as luminance increases from 0.025 cd/m^2 to 60 cd/m^2, above 80 cd/m^2 the increase is very slight, and above 500 cd/m^2 it is negligible[158]. Visual acuity therefore remains relatively independent of luminance over a wide range of photopic luminances

extending from the equivalent of full moonlight to a bright sky on a sunny day[159]. However, it should be noted that pupil size will varies with luminance such that the influence of luminance is likely to be greater in the presence of uncorrected refractive error. Notwithstanding this caveat, the fact that visual acuity is relatively independent of luminance is helpful in view of the fact that acuity measurement may be carried out with a number of display modalities including retro-illuminated light boxes, projection charts, and computer monitors. It may be difficult to achieve a specific display luminance with each of these types of displays. Recommendations for visual acuity chart luminance have ranged from 85 to 300 cd/m^2 according to Bailey[3] who has advocated a clinical tolerance of e.g. 80 to 320 cd/m^2 as reasonable and practical[3]. In view of the relationship between luminance and visual acuity (see Fig 2.10.8) it would appear that the upper limit of this range is of lesser importance than the lower limit.

Figure 2.10.8. The relationship between visual acuity and luminance after Shlaer[159]



80 cd/m2 equates to +3 log Trolands (for a pupil diameter of 4mm)

As stated in section 2.5.6, spatial resolution improves with increasing contrast. At a contrast of 80% visual acuity is approaching a maximum[160], and the contrast of conventional visual acuity charts typically exceeds this level[3]. Acuity measurements carried out at lower levels of contrast are measuring a different aspect of visual

function from that measured at high contrast, and the former are often used to compliment the latter (see section 2.5.6).

2.10.10. *Measurement time*

Historically, the time required for an acuity measurement has received little attention in the scientific literature. When considering this issue, it should be noted that the letters close to a subject's threshold are the ones which will tend to be most time consuming to attempt. Conversely, a series of letters which can be easily resolved will typically be read fluently and swiftly. Contemporary recommendations on acuity testing advocate the use of interpolated scoring with a forced choice methodology, and strict termination rules[50 51]. This approach will increase the number of near-threshold letters which the subject is required to attempt, thereby increasing test time. It is likely that in routine clinical practice, a less rigorous approach to acuity measurement has resulted in short measurement durations. The move towards a more rigorous approach to acuity testing may require more consideration of measurement duration. It may also explain the reluctance to adopt contemporary logMAR charts in routine practice.

2.10.11. *Range of letter sizes*

Snellen's original chart was used at a distance of 20 feet and covered an acuity range of 6/60 to 6/6 (+1.0 to 0.0 logMAR). The use of a minimum stimulus size which is well within the capabilities of many normal individuals will result in a truncated distribution of measured acuity (see sections 2.11 and 2.12). The implications of this were understood by Green[95] who accordingly utilised a minimum letter size equivalent to Snellen 6/3 (-0.30 logMAR) in his visual acuity chart. More recently, those contemporary logMAR charts produced by Sloan[98], Bailey & Lovie[92], and Ferris[94] all feature a minimum stimulus size which is designed to avoid a truncated distribution. The majority of visual acuity chart designs feature a maximum stimulus size which equates to that of the Snellen chart (6/60 or +1.00 logMAR). This provides a range which is sufficient to cover a substantial proportion of those acuities encountered in practice. The range of acuities measurable using a visual acuity chart may be extended through manipulation of the distance from which the chart is viewed. In the absence of

a carefully chosen progression of letter sizes this may cause difficulties relating to the measurement scale, as discussed in section 2.12.

2.10.12. *Viewing distance*

Implicit in the specification of visual acuity in terms of the minimum angle of resolution (see section 2.10.5) is the assumption that visual acuity is independent of the viewing distance. It seems natural to assume that providing the retinal image size created by an acuity stimulus is constant, the distance of the stimulus from the observer will be irrelevant. This hypothesis would seem to be supported by some of the early investigations into the relationship between visual acuity and testing distance which suggested that, for distances in excess of 2 metres, visual acuity was independent of the testing distance[161 162]. However, there are a number of factors which, at least in theory, might influence the relationship between viewing distance and visual acuity. Firstly, any uncorrected refractive error (and in particular myopia) will result in a clearer retinal image as the chart approaches the subject's far point. Also, even in the absence of refractive error, a subject viewing a chart from a distance of 2 metres needs to be exert 0.5 dioptres of accommodation to maintain a clear retinal image. An appropriate refractive correction will be required for subjects who are unable to sustain the required degree of accommodation. Even if the subject possesses sufficient accommodation, a reduced working distance may result in a smaller pupil size (either through accommodative miosis, or through increased retinal illuminance). In theory, for pupil sizes of 3 mm and larger miosis might be expected to improve visual acuity through the reduction of aberrations (see section 2.4.1). This effect would be much more pronounced in the presence of refractive error.

The scientific literature contains conflicting evidence as to the relationship between viewing distance and visual acuity where the distance is less then 2 metres. Some investigators have found that testing distances of less than around 2 metres are associated with poorer acuity measurements[161 163-165], whereas others have shown an increase in acuity for close testing distances prompting some workers to doubt the rigor of the scientific methods employed in these studies[125 166]. Recent work in this area appears to suggest that testing distances of less than 2 metres are indeed associated with poorer acuity measurements, even after consideration of the degree of

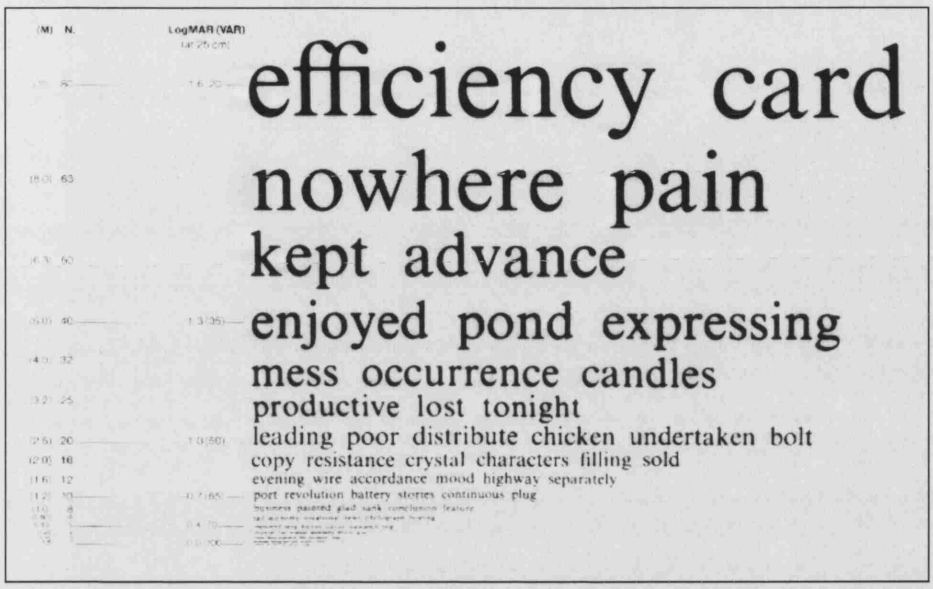
refractive error and the required accommodation[167]. It is significant that this study utilised contemporary letter charts which remove some factors which may have the potential to confound Snellen acuities measured at different distances (e.g. although a Snellen acuity of 6/12 at measured at 1.2 metres subtends the same angle as a 6/60 letter at 6 metres, the task is very different in terms of the number of letters on the line and the degree of contour interaction – see figure 2.10.7 and section 2.10.8).

Acuity charts are typically used at 4 or 6 metres and there is currently no evidence to suggest that acuities measured at these two distances will not be comparable. The existing evidence does suggest however that studies utilising testing distances of less than 3 metres to measure acuity (as is often required for poor acuities) should be aware of this potential source of bias.

2.10.13. *The measurement of near acuity*

If the resolving power of the eye is considered in angular terms, then (after refractive considerations) the distance at which the measurement is conducted might be considered irrelevant. Indeed, tests of near acuity exist which are scaled down versions of letter charts for the measurement of distance acuity (e.g. those of Ferris[94] and Ricci et al[160]). These tests are designed to produce results which are comparable to those of their distant counterparts. More common however, are tests of reading ability such as those of Bailey[168] (see Fig 2.10.9) and Wolffsohn[169]. Tests of this type tend to produce poorer estimates of visual performance than might be predicted from a distance visual acuity measurement[170]. This occurs because, although dependant on visual acuity, reading is a more complex task influenced by various other parameters[171], and certain ocular conditions are associated with a reduction in visual performance which is more pronounced for more complex tasks[138]. With the exception of the concept of using distant visual acuity charts at varying distances, this thesis will not consider further the measurement of acuity at near.

Figure 2.10.9. The Bailey-Lovie word reading chart



2.10.14. *The measurement of acuity in children*

The act of undergoing a standard distance visual acuity measurement constitutes a complex cognitive task. The stimulus first of all must be resolved and interpreted. This interpretation then has to be compared with a series of possible alternatives from prior experience, and once selected the most likely match has to be communicated to the examiner. As such, the use of standard distance acuity charts is not appropriate for all age groups and mental abilities. A variety of visual acuity tests have been designed for use with children[172 173] and others[174-176] who are unable to cope with the standard tests. These range from tests which rely on visually evoked potentials (see section 2.10.14), through to those for which an unwitting response is elicited (e.g. preferential looking tests), to simplified versions of adult tests in which the subject is asked to name the stimulus, or to match it with a suprathreshold example of the stimulus. The range of available tests can be considered in a hierarchy (see Table 2.10.3) from which the highest test which is within the capabilities of the child is selected[177]. It is advised that acuity should be measured using standard distance tests as soon as the child is capable of complying with such tests[3]. The following categories of paediatric acuity tests are listed in increasing order of 'difficulty'.

Table 2.10.3. The hierarchy of children's visual acuity tests.

Test	Description	Example
Fully objective tests	See section 2.10.14	VEPs, OKN
Preferential looking	Based on the assumption that a child will be attracted towards the detail in a stimulus if s/he can resolve it	Teller & Cardiff acuity cards
Single-symbol cards	Measures the smallest symbol which can be named	Kays pictures
Single-letter cards	As below but often using letters	Sheridan-Gardener
Multiple-letter cards	As below but more than one stimulus displayed at a time	Glasgow acuity cards
Letter charts	Standard charts but using symbols or letter sets designed for children	Lea symbols

Listed in increasing order of difficulty

Visually Evoked Potentials (VEP) and Optokinetic Nystagmus

Fully objective tests for those with very poor responses (see section 2.5.5).

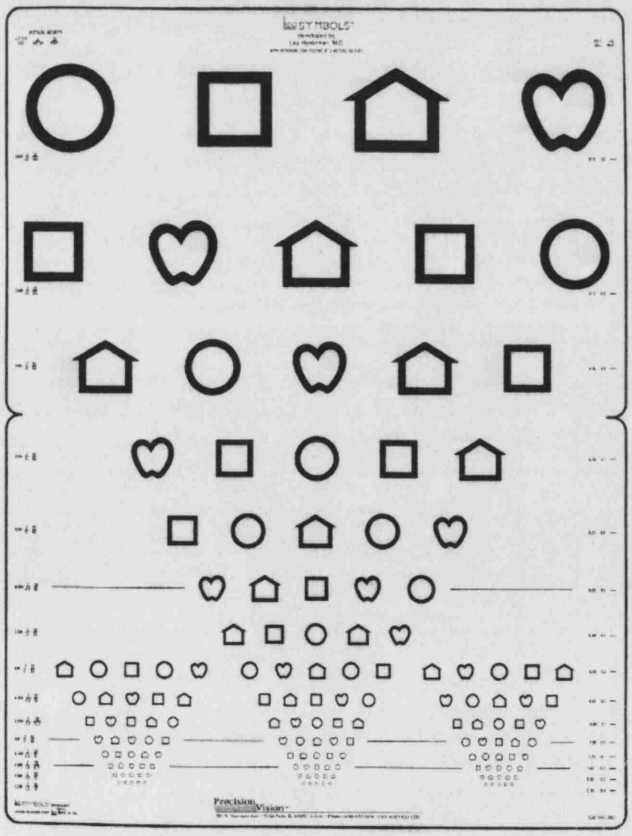
Preferential looking

These tests rely on the fact that a child will be attracted towards a stimulus which contains detail providing they can resolve the detail. The Teller preferential looking cards[178] feature two target areas, one of which is contains a grating of a certain spatial frequency, the other contains a uniform grey field. If the grating can be resolved, the child will be attracted to it. If it cannot be resolved, it will appear indistinguishable from the grey field as its space averaged luminance is the same. The examiner observes the child from behind the card and is masked to which side of the card the stimulus is positioned. This is intended to avoid bias on the part of the examiner when deciding whether the child is attracted to one side of the card more than the other. The examiner judges the highest resolution grating which attracts the child's attention. The Cardiff cards[179] are based upon the same principle but feature outlines of familiar objects drawn in a white line flanked by a pair of black lines. The object will only be visible if the lines can be resolved.

'Flashcard' tests

The subject is presented with a series of progressively smaller stimuli. The stimuli may be letters[180] or symbols[181]. These are either named, or matched with those on a separate card featuring large examples of each of the stimuli. The test is typically introduced as a game to promote the child's interest. The traditional version of this test is the Sheridan-Gardiner test which features isolated letter optotypes[180]. The design of such tests has been of concern because the visual deficits in amblyopia (the detection of which is the common reason for using such tests) are underestimated in the absence of adjacent contours[148 151]. This has led to the development of tests such as the Glasgow acuity cards[182] which incorporate 'crowding bars' as well as other benefits of contemporary chart designs. The Glasgow acuity cards have been shown to be more sensitive to amblyopia than the Sheridan-Gardener format[183 184], and agree well with the current gold standard visual acuity test[184].

Figure 2.10.10. Chart based upon the Bailey-Lovie design but featuring LEA symbols



Children's letter charts

Versions of contemporary letter charts are now available with letters or symbols which are designed specifically for children[185] (see Fig 2.10.10). Such charts are most comparable to the standard tests in routine use. These tests mark the last step in the progression towards measuring a child's acuity on standard acuity charts.

2.10.15. *Computerised measurement of acuity*

Computers have been employed for around 40 years as a method of displaying stimuli in vision research. The scientific literature contains numerous papers describing computerised visual acuity tests designed for clinical use[186-193]. Some tests simply use the combination of computer and monitor to create a highly versatile display, whilst others introduce a degree of automation into the presentation of stimuli and the collection of subject responses. The potential advantages of computerised visual acuity tests include:

- Increased choice of stimuli,
- Pseudo-random ordering of stimuli,
- Fewer examiner-related sources of bias,
- Variable contrast,
- Automatic calculation, storage and analysis of acuity score,
- Generation of psychometric functions,
- Variable testing distance,
- Ability to display one stimulus at a time
- Ability to display very large stimuli, and
- Partial credit for stimuli which are narrowly misnamed[194].

Another potential advantage of a computerised acuity test relates to the ability to present randomly generated stimuli and to record responses. This allows for visual acuity to be measured repeatedly and the results analysed. Brown and Lovie-Kitchin have suggested that this may allow an individual's own degree of test-retest variability to be established. In turn this may be used to detect change earlier than would have been possible using an index of variability derived from a group of subjects, some of whom may demonstrate a high degree of variability[143 195].

There are some potential disadvantages to computerised visual acuity testing which include:

- A 'warm-up period' before luminance stabilises,
- Lower luminance then conventional retro-illuminated chart displays,
- Flicker due to interaction between small eye movements and a 'raster-scan' refresh mechanism, and
- Increased cost.

However, the first three of these caveats have been negated through the development of flat screen displays, whilst the last is becoming less of a drawback with the progressively reducing cost of computer hardware.

2.11. The Optimum Visual Acuity Chart Design

The perfect acuity test would be objective, instantaneous, unbreakable, take up little room, be easily transportable, and affordable. It would produce measurements on any subject which are accurate, precise, and readily analysable. Although many of these qualities are beyond the reach of any visual acuity chart, it is possible to speculate, based upon our existing knowledge of the relationship between the design and performance of visual acuity tests, on the optimum design for a letter chart. The design of existing charts can then be critically assessed in the light of this information.

Optotype legibility

If a chart contains more than one stimuli at each size level, then these stimuli should be equally legible, such that the difficulty of a stimulus is determined purely by its size. The presence of variable legibility may introduce confounding effects.

Number of optotypes per line

There should be the same number of optotypes at each size level. This helps to eliminate variables over and above letter size. For example, with respect to contour interaction, if 5 letters are used on each line, then each line has three letters which are crowded from both sides, and two which are crowded from only one side. Also, the number of letters per line and the inter-line size increment dictates the test's scale increment which should be sufficiently fine to achieve adequate precision (see below).

Progression of optotype sizes

The size interval between adjacent lines on the chart should be consistent across the chart. The choice of size progression should also be such that a just-noticeable difference represents the same number of scale units across the chart. This is best achieved using a geometric progression of letter sizes[130].

Scale increment

An excessively coarse scale is known to be associated with reduced test precision (see section 2.10.6). The scale increment should be sufficiently fine to allow adequate precision. The use of fixed number of letters per line combined with a geometric progression of letter sizes, facilitates the use of an interpolated scoring method which involves allocating each letter on the chart an equal value (see section 2.10.6). In this

way credit can be given for each letter correctly read, resulting in a finer scale increment than would be dictated by the size interval between rows of letters.

Range of optotype sizes

The range of optotype sizes should be sufficient to allow a wide range of acuities to be measured. The range of stimulus sizes should exceed the range of acuities likely to be encountered such that there is no truncation of the distribution of measured acuities. Collectively the number of optotypes per line, size interval, and range of optotype sizes determines the time required for a measurement.

Optotype spacing

The vertical and horizontal spacing between optotypes should be consistent relative to optotype size across the range of acuities to avoid variable contour interaction as a potential confounding variable.

Notation

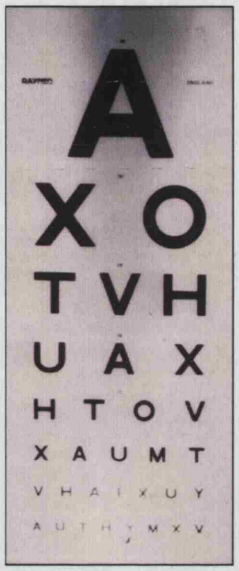
The acuity notation should be of a form that is amenable to the use of interpolated acuity scoring and statistical analysis.

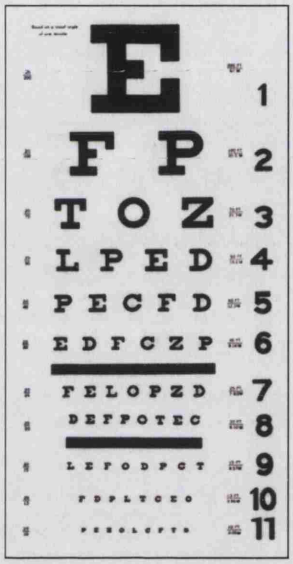
Measurement time

The length of time required for a measurement should be such that the use of the chart is not precluded in the particular environment in which it is to be used. For example, it may be that a test which requires a longer time for a measurement can be tolerated in a prospective clinical research trial, if that test offers some other advantage in terms of performance. In a routine clinic environment, there may be more constraints on time such that a pragmatic compromise trading off performance against test duration is preferable.

2.12. The Snellen Chart Design

Figure 2.12.1. Two sample Snellen charts.

A) 

B) 

Optotype legibility

The Snellen chart uses letter stimuli. The difficulty of the letters of the Roman alphabet is known to vary (see section 2.10.2) and the design of the Snellen chart does not control for this factor. As such, the change in stimulus difficulty going, for example from a letter on the 6/18 line to a letter on the adjacent 6/12 line, may not be proportional to the change in size. Also, the difficulty of the 6/18 line on one version of the chart may be different to the difficulty of the 6/18 line on another version of the chart which features different letter combinations.

Number of optotypes per line

The number of letters per line on a Snellen chart varies between 1 and 8. Accordingly, the proportion of letters which are subject to contour interaction from both sides varies from none (for the 6/60 line) to 75% (for those rows featuring 8 letters).

Progression of optotype sizes

The Snellen chart features an irregular progression of letter sizes (see Table 2.12.1), the largest size interval being more than three times the smallest. This has implications for measurement precision which is influenced by the inter-line size increment, and will accordingly vary according to where on the chart the measurement is terminated. In addition to this, the progression of letter sizes differs from one version of the chart

to the next, introducing a further source of variability. This irregular size progression invalidates the specification of change in terms of the number of lines (or letters) because e.g. a two line improvement from 6/60 cannot be considered equivalent to a two line improvement from 6/9. An additional disadvantage of an irregular progression of letter sizes is that changes in acuity of a given number of lines measured at one testing distance, cannot be equated to the same change in acuity measured at a different distance. For example, for a testing distance of 6 metres, a reduction in acuity of one line from 6/36 is 6/60; whereas at 3 metres, one step worse than 3/18 (which is equivalent to 6/36) is 3/24 (which is equivalent to 6/48).

Table 2.12.1. Stimulus size increments of Snellen and ETDRS chart designs

Snellen			ETDRS	
Snellen	**logMAR equivalent**	**Increment***	**logMAR**	**Increment***
6/60	1.00	1.66	1.00	1.26
6/36	0.78	1.51	0.90	1.26
6/24	0.60	1.32	0.80	1.26
6/18	0.48	1.51	0.70	1.26
6/12	0.30	1.32	0.60	1.26
6/9	0.18	1.51	0.50	1.26
6/6	0	1.20	0.40	1.26
6/5	-0.08	-	0.30	1.26
			0.20	1.26
			0.10	1.26
			0	1.26
			-0.10	1.26
			-0.20	1.26
			-0.30	-

* – inter-line increment as a multiple of the line below

Scale increment

The variable number of letters per line combined with the variable scale increment prevents the use of an interpolated scoring method. Because the letter values on the Snellen chart vary from one line to the next, a time consuming calculation would be required to produce an interpolated score. Even if the time is taken to use an interpolated scoring method, the variation in scale interval across the chart will be considerable, with the largest scale increment being more than 20 times the size of the smallest. As already discussed, this will result in a considerable variation in test precision according to whether the threshold falls near the bottom (as with good acuities) or near the top (as with poorer acuities) of the chart. The irregular scale increment also complicates the measurement of visual acuity at non-standard testing distances, as the size of scale increment is dependent upon both the underlying level of acuity and the viewing distance.

Range of optotype sizes

The original Snellen chart design (see Fig 2.10.2) features letter sizes ranging from 6/60 (equivalent to +1.0 logMAR) to 6/6 (equivalent to 0.00 logMAR). This represents a considerable truncation of the scale in view of the fact that for normal young adults, *average* acuity has been shown to be -0.16 logMAR (equivalent to Snellen 6/4)[142]. Although it is debatable whether Snellen designed his chart this way by design or in error, the effect is to produce an incomplete distribution of acuity. This is important from a statistical, as well as a clinical point of view. Clinically, truncation of the measurement scale may mask changes in acuity around the upper end of the normal range. For example, a subject whose true acuity is 6/4 reading a chart on which the smallest stimulus size is 6/6, will achieve an acuity score of 6/6. Should his true acuity deteriorate from 6/4 to 6/6, he will still achieve an acuity of 6/6 when his acuity is retested, and the change in acuity will go unidentified. From a statistical point of view, an incomplete distribution can have a significant impact upon the evaluation of test performance. If we again consider the example of the chart whose smallest stimulus size is 6/6, a comparison of test precision between a group with very good acuity and one with abnormal vision, will suffer from bias as the acuity score of the normal group is restrained by the truncation. The variation will remain undetected as all the stimuli are suprathreshold for both test and retest. Therefore the precision estimate for the

normal group will be artificially inflated. At 6/60 (+1.00 logMAR), the largest stimulus size is sufficiently large to cover a significant proportion of the measurements encountered in clinical practice. However, the measurement of very poor acuities (for example in low-vision practice) requires manipulation of viewing distance. The use of the Snellen chart at varying distances is complicated by the variable inter-lien size interval as discussed under 'progression of optotype sizes'.

Optotype spacing

The horizontal spacing between letters on the Snellen chart measured in terms of multiples of the width of the letters on that line varies from one line to another (see Fig 2.10.2). For example, one common design of Snellen chart (there being several versions with varying numbers of letters, and spacing between rows/letters) features horizontal spacing which varies between 42% and 114% of the width of the letters on the respective line. There is also no consistent relationship between the vertical spacing between lines and letter height. For example, on the Snellen chart version shown in Fig 2.12.1A, there is a 25mm vertical space between each row of letters on the chart. This means that expressed in multiples of the height of the letters in the lower row, the 6/5 and 6/6 rows are separated by approximately 3.5 letter heights, whereas the 6/36 and 6/60 rows are separated by only approximately 0.5 letter heights.

Notation

The acuity notation used in the Snellen chart is the Snellen fraction (section 2.10.5) e.g. 6/12 (+0.30 logMAR). The Snellen fraction is not amenable to statistical analysis. As previously discussed, despite the frequency with which it is done, to speak of a change of e.g. 2 lines of letters is inadvisable, since the corresponding change in resolving power depends upon the region of the chart in which the change has occurred. Although the Snellen fraction may be converted to decimal form (e.g. 6/12 becomes 0.5), because of the irregular progression of letter sizes, this cannot be considered interval data (see section 2.10.5) which renders the use of basic arithmetic operations invalid[135]. The statistical analysis of Snellen visual acuity data should therefore be limited to less powerful non-parametric procedures. A form of interpolation is often attempted with the Snellen chart (e.g. 6/12+2). However, although this may be of some use clinically, this form of notation is also meaningless from a statistical perspective.

Measurement time

The time taken for a Snellen acuity measurement has not been addressed in the literature. In clinical practice Snellen acuity measurements are often carried out quite quickly. However, a measurement produced in this way is likely to be subject to various sources of bias (see section 2.10.7). For example, the subject is often asked to decide for themselves the smallest letter size which they can manage. This may result in an underestimate of acuity which is more pronounced for timid subjects than for bold ones who are happy to push themselves further. It is rare to see Snellen acuities conducted using time consuming paradigms which feature forced choices, and strict endpoint definitions. In general, the approach which allows a Snellen measurement to be conducted quickly, is also one which is likely to be subject to various sources of bias.

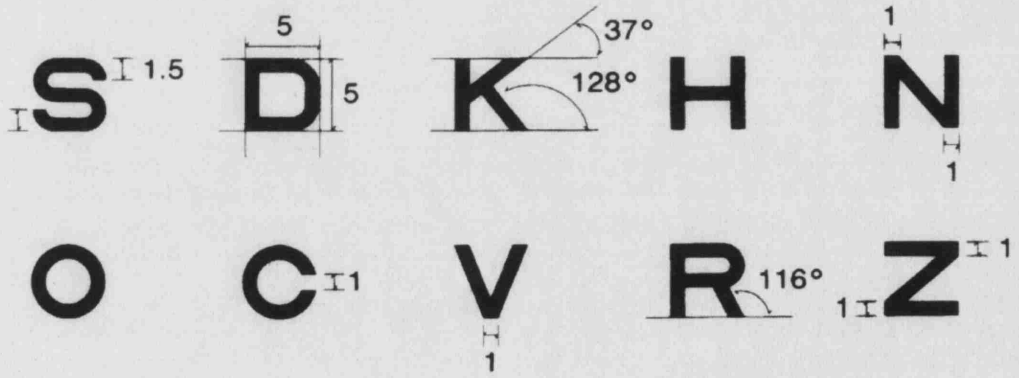
2.13. The Bailey-Lovie & ETDRS Charts

The Early Treatment Diabetic Retinopathy Study (ETDRS) logMAR chart[94], along with the Bailey-Lovie chart[92] on which it was based are currently considered the gold standard method of measuring acuity in clinical research. They represent an attempt to eliminate extraneous variables such that the difficulty of each line on the chart is influenced by letter size only. There are three chart versions in total: charts 1, 2, and R. The principles described in this section are those to which the design of charts 1 and 2 adhere. The R chart stands for refraction and differs only from the other two charts in that it lacks the rigorous control of letter difficulty which is incorporated in the design of charts 1 and 2.

Optotype legibility

The ETDRS chart uses a set of 10 letters advocated by Sloan as suitable for the purposes of visual acuity testing[98 101] (Fig 2.13.1). Each letter is designed on a 5x5 grid and lacks serifs. The letters were chosen such that their legibility is approximately equal, and such that their average legibility equates to that of the Landolt C[101] (see section 2.10.2).

Figure 2.13.1. The Sloan set of approximately equally legible letters[98]



Although the letters in the Sloan set are chosen such that legibility is approximately equal, there is a degree of variability within the set. Table 2.13.1 lists the 10 letters in decreasing order of legibility. The range is such that on average, the easiest three

letters (Z, N, and H) are more than 20% easier to read than the three most difficult (C, O, S). To minimise this potential source of extraneous variability, the letter combinations for each line of the ETDRS chart were chosen such that the total difficulty of each line is approximately equal. The Bailey-Lovie chart follows the same principle but uses a different set of 10 letters adopted by the British Standards Institution[49] (D, E, F, H, N, P, R, U, V, and Z). These letters happen to be based upon a 5x4 grid rather than the 5x5 grid of the ETDRS chart, resulting in a more horizontally compact chart (Fig 2.10.5).

Table 2.13.1. Relative difficulties of the letters in the Sloan set after Sloan[98]

Letter	% of correct responses at threshold	Deviation from average % correct
Z	94.0	+12.0
N	91.6	+9.6
H	89.3	+7.3
R	86.3	+4.3
V	84.6	+2.3
K	82.1	+0.1
D	79.5	-0.5
C	71.4	-10.6
O	71.0	-11.0
S	70.6	-11.4
Average	**82.0**	

Number of optotypes per line

Both the ETDRS and Bailey-Lovie charts feature 5 letters on each row of the chart. This contributed towards task consistency across the chart.

Progression of optotype sizes

Each row on the Bailey-Lovie and ETDRS charts features a geometric progression of letter sizes such that the letters in each row are 1.26 times larger than the row beneath. This equates to a size interval of 0.1 log units. Such a progression has been shown to produce an equal discriminability scale (see section 2.10.4.).

Scale increment

As mentioned earlier, the size interval is 0.1 log units. Hence if the chart is scored using a line assignment method, the scale interval is 0.1 log units. However, using a method originally suggested by Kitchin and Bailey[138] the inter-line interval is divided by the number of letters per line such that each letter on the chart is allocated a value of 0.02 log units. This results in a finer scale increment of 0.02 log units which is known to be associated with improved precision[78 80] (see sections 2.10.4 and 2.10.6). It is also advantageous that the size of the scale increment is independent of scoring distance.

Range of optotype sizes

The Bailey-Lovie and ETDRS charts feature stimulus sizes which cover the range +1.00 to -0.30 logMAR (Snellen equivalent 6/60 to 6/3). There is evidence to suggest that extending the stimulus range to -0.30 logMAR is unlikely to result in a truncated distribution of acuity even of young normal subjects[142]. With respect to the opposite end of the scale, the use of a logarithmic progression of letter sizes allows acuities poorer than +1.00 logMAR (Snellen 6/60) to be measured through manipulation of the chart viewing distance without altering the scale increment. Scoring Bailey-Lovie and ETDRS acuities at varying testing distances is straightforward as a simple formula may be used to correct the acuity score for any non-standard viewing distance e.g.

$$VA_C = VA_M + \log10(d/4)$$

Where VA_C is the distance-corrected acuity score, VA_M is the measured acuity score, d is the viewing distance at which the acuity measurement was taken and 4 metres is the standard viewing distance[A]. The ease with which acuity scores can be adjusted for non-standard viewing distances allows acuity to be measured over a wide range without the need for multiple large stimuli which in turn increase the size of the chart.

Optotype spacing

The ETDRS and Bailey-Lovie charts both maintain consistent spacing between letters and lines across the chart. On each line of the chart, the spacing between letters is equal to the width of letters on that line. Letter size on both charts is specified

[A] Therefore for viewing distances twice the standard, add 0.3 logMAR to the measured acuity score, and for a viewing distance half the standard subtract 0.3 logMAR from the measured acuity score.

according to the vertical dimensions of the letter, such that because the letters on the Bailey-Lovie chart are based on a 5x4 rather than a 5x5 grid, each letter (and accordingly, each space between letters) on that chart is 20% narrower than on the ETDRS chart. With respect to vertical spacing, the two charts are identical, both featuring rows which are separated by a space equal to the height of the letters in the lower row.

Notation

The Bailey-Lovie and ETDRS charts use logMAR notation. This can be considered interval data and accordingly is suitable for analysis using parametric statistical methods (see section 2.10.7).

Measurement time

Although it has been suggested that the time required for an ETDRS measurement is '...probably one of the main reasons limiting widespread acceptance of ... ETDRS charts in everyday practice...'[196], the subject has received little attention in the scientific literature. It is likely that the time required for a Bailey-Lovie or ETDRS measurement will be greater than that required for a typical Snellen measurement. This is due in the main to the use of forced choice methods and rigorous termination rules with these charts, which are never employed with the Snellen chart (see sections 2.10.7 & 2.10.10).

3. SUMMARY & AIMS

Summary

- Visual acuity is a measure of the spatial resolving power of the visual system.
- Visual acuity is the most important measure of visual function in ophthalmic and optometric clinical practice as well as clinical research.
- Letter charts are widely accepted as the method of choice for measuring visual acuity in subjects who are capable of undergoing such a measurement.
- The performance of a visual acuity chart is determined by both the chart's design and the method of use.
- Snellen's chart has a flawed design, and the method with which it is typically used is prone to various sources of bias.
- The design principles upon which visual acuity charts are based have become increasingly robust culminating in the Bailey-Lovie and ETDRS logMAR charts which have become the gold standard in clinical research.
- The uptake of these logMAR charts in clinical practice has been negligible. This may be due to one or more of the following reasons:
    - They are perceived as time consuming
    - They use unfamiliar notation
    - There is a lack of awareness of the advantages of logMAR charts and/or the Snellen chart's limitations
    - Cost

Aims

Initial aim

- To develop a visual acuity chart which is based upon rigorous design principles, and which allows accurate acuity measurements to be taken in a clinically acceptable period of time, with greater precision than the Snellen chart.

Subsequent aims

- To develop a test capable of conducting repeated measurements of visual acuity and calculating their average (the 'PC-test').
- To determine whether PC-test acuities are accurate compared with ETDRS acuities, and whether they are more precise.
- To investigate whether a relationship between optical defocus and the precision of visual acuity measurements exists in normal subjects.
- To determine whether the intra-test variability within a PC-test can be used to create an individualised change-criterion and hence improve the detection of change.
- To assess the performance of the ETDRS chart and the PC-test in terms of specificity and sensitivity to change.
- To develop a simple mathematical model to predict sensitivity to change from a knowledge of test precision.

4. EXPERIMENTAL DESIGN AND ANALYSIS

Certain aspects of experimental methodology will be employed repeatedly throughout this thesis and hence will be described in detail here for future reference.

4.1. THE ANALYSIS OF METHOD COMPARISON STUDIES

Where a new clinical test is being proposed as a potential replacement for an existing test, an evaluation of the performance of the new test is required. The new test will typically offer some advantage over the existing test such as reduced cost, bulk or measurement time. Where known quantities can be measured using the new test, the process is one of straight forward calibration. When this is not possible (as is often the case), the existing test is often adopted as a reference standard, and the performance of the new test compared with that of the reference standard. It is advantageous for the results of a new test to agree well with those of the existing test, otherwise a new normative range must be established and the monitoring of disease over time will be compromised. There are a number of statistical approaches for comparing the performance of a new test with that of a reference standard test. A common form of analysis used in such 'method comparison' studies is correlation. The correlation coefficient is an index of the strength of the relation between two variables, and is generally seen in one of two forms. The 'product moment correlation coefficient' (sometimes referred to as the 'Pearson product moment correlation co-efficient', but generally shortened to just 'correlation coefficient') allows the relation between two variables to be evaluated whether or not they are expressed in the same units. The less common 'intra-class correlation coefficient' or ICC is a variant of the standard correlation coefficient, and is used where genuine replicates are being considered (i.e. in situations where the order of the observations to be compared is unimportant). The statistic 'r' is used for both forms of correlation coefficient and will vary between -1 and +1. These extremes of the scale indicate a perfect linear relation (with negative and positive gradients respectively), whereas r=0 indicates no relation at all. Numerous method comparison studies have used one form of correlation or another, including many studies of visual acuity[119 196-199].

Bland and Altman[200-202] have suggested that correlation is inappropriate for analysis of method comparison studies for a number of reasons, some of which are pertinent to the

analysis of visual acuity data. Firstly, although correlation assesses the strength of a relation between two variables, it does not necessarily follow that this is synonymous with agreement. For example, if a test produces measurements which are exactly half those of another test, the two will be perfectly correlated i.e. their measurements could be plotted against one another producing a straight line (r=1). The level of agreement, however, would be poor, as shown by the fact that they could not be used interchangeably. Also correlation is influenced by the range of data in the sample upon which the analysis is based. A large range of values will ensure a high level of correlation even in the presence of a considerable lack of agreement. With respect to the analysis of acuity data, it may be argued that this will not affect a comparison of two tests on a given series of individuals, but it makes any comparison across studies difficult.

Bland and Altman[200-202] have proposed an alternative method of assessing agreement between two methods of clinical measurement. Using their approach, agreement is considered to be the extent to which the results of one test could be substituted for those of the other, rather than considering solely whether a relation exists. A potential disadvantage of the Bland-Altman method is that accuracy is expressed in the same units as the original measurements. This dictates that both measurements must be in the same form of notation. However, a significant benefit of this fact is that the results can be readily related back to the units in which the original measurements were taken, which facilitates interpretation. For example, a mean difference between the results of two tests of +0.10 logMAR suggests that, on average, one test 'reads' one line of letters better than the other. The interpretation of a correlation of 0.87 between two tests is less intuitive. Before considering the Bland-Altman approach to agreement, it is preferable to consider their approach to two fundamental aspects of test performance which together determine the level of agreement: accuracy and precision.

4.1.1. *Accuracy*

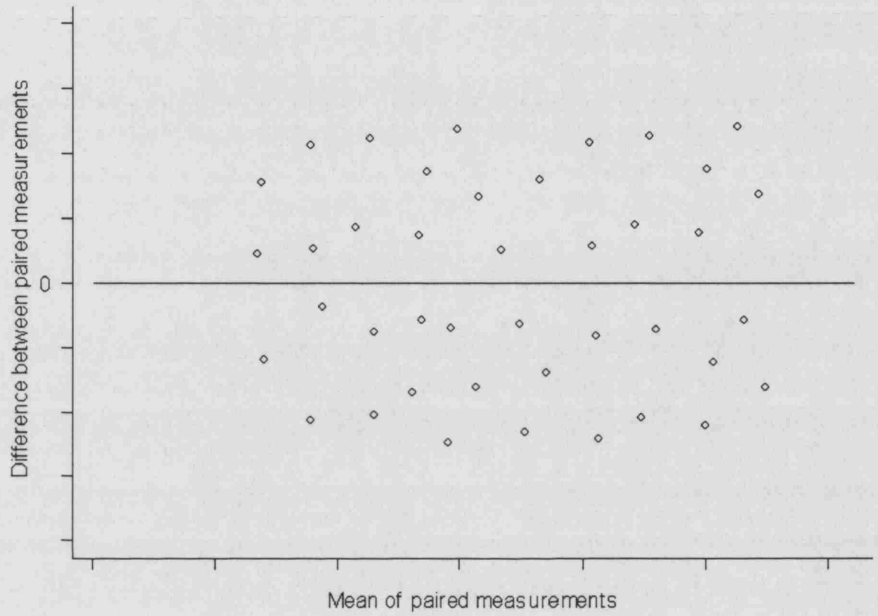
Accuracy is defined as the degree to which, on average, an estimate represents the true value of what is being measured. Accordingly, an accurate test is one whose measurements (on average) show no bias when compared with the true value (or a

reference standard where the true value is unknown). Where a new test is proposed to replace an existing test, it is desirable that the new test produces accurate data. To quantify accuracy, measurements are performed on a sample of subjects using both the reference standard test and the new test, creating a sample of paired measurements. Rather than assessing the correlation between these paired measurements, the Bland-Altman method requires that for each subject, the result of the new test is subtracted from that of the reference test producing a series of 'differences'. The spread of these differences is largely determined by the combined measurement errors of the two tests, and will therefore tend to conform to a normal distribution (regardless of whether or not the measurements themselves are normally distributed). If the new test is accurate when compared with the reference test, then the mean of the differences will be zero. Any departure from zero indicates inaccuracy (i.e. a degree of systematic bias). The importance of any departure of the mean from zero can be determined statistically by calculating the 95% confidence interval of the mean. Should the 95% confidence interval not include zero, it can be said that a notable degree of inaccuracy exists[203]. Such a finding should however be interpreted in the light of the smallest clinically important degree of inaccuracy. If a comparison of two tests yields a mean difference which is of a clinically unimportant magnitude, then this may not be an obstacle to using the tests interchangeably even if the 95% confidence interval for the mean excludes zero. Conversely, if the number of subjects in the sample is small, the size of the mean difference may be clinically important but the 95% confidence interval may still be sufficiently wide to include zero.

A graphical representation of degree of accuracy is possible via the use of a 'Bland-Altman plot'. A Bland-Altman plot is a scatter plot in which the difference between the measurements taken on the two tests is plotted against their mean. Fig 4.1.1 shows a sample Bland-Altman plot. The further a point from the horizontal midline, the larger the difference between the two test results for that particular subject. Points towards the left of the chart relate to subjects with a good underlying level of acuity, whereas those to the right relate to poorer levels of vision. The points in Fig 4.1.1 are evenly distributed above and below the horizontal midline. This suggests that, on average, the new test is accurate compared with the reference standard test, and the mean of the differences will be close to zero. The horizontal spread of points in Fig 4.1.1 indicates

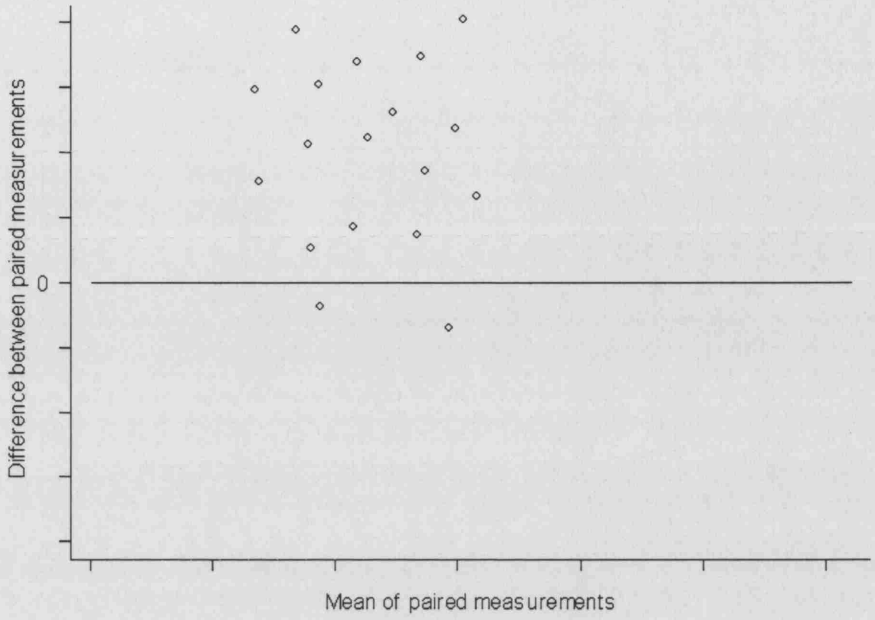
that the subjects on whom these measurements were taken exhibited a range of acuities. Had the subjects had more similar acuities, the horizontal grouping of points would have been much tighter.

Figure 4.1.1. Bland-Altman plot showing a high degree of accuracy



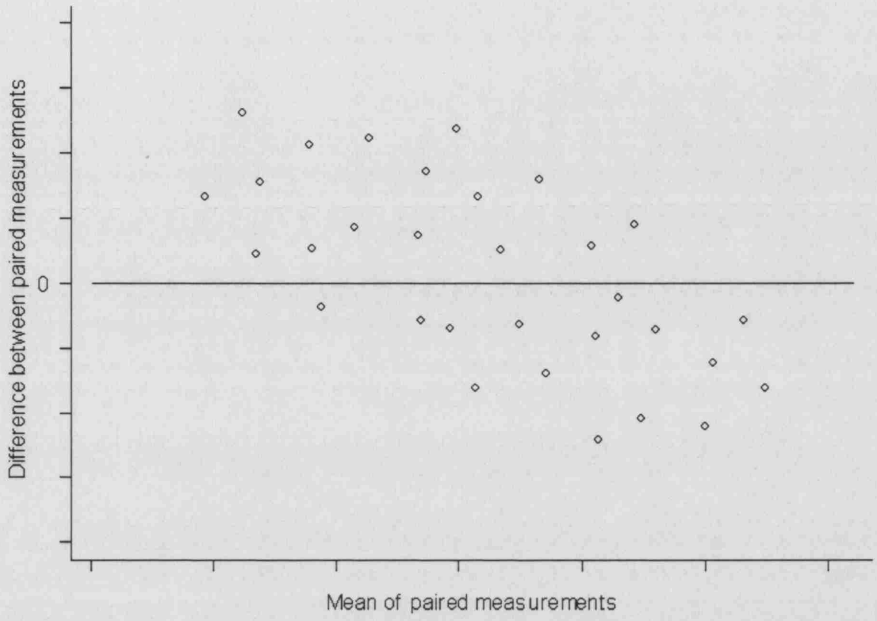
In contrast to the appearance of Fig 4.1.1, the vast majority of points in Fig 4.1.2 are situated above the horizontal midline. This indicates that acuities measured with the new test are subject to a degree of systematic bias when compared with those of the ETDRS chart. It is also evident from this plot that the range of underlying acuity is not as wide as that featured in Fig 4.1.1.

Figure 4.1.2. Bland-Altman plot showing a low degree of accuracy



An important use of the Bland-Altman plot, in addition to showing the range over which the subjects acuities fell, relates to acuity-related bias. The Bland-Altman plot shown in Fig 4.1.3 has an approximately equal number of points above and below the horizontal midline. The mean difference for these points would be approximately zero which is suggestive of a high degree of accuracy. However, closer inspection reveals a tendency for those points towards the left of the plot to fall above the line, whereas those towards the right end of the plot are mostly below. This shows that although the new test is *on average* accurate when compared with the reference test, the new test tends to overestimate the measurement at one end of the range, and underestimate it at the other end. In other words the degree of accuracy varies according to the underlying level of acuity. Thus, the use of Bland-Altman plots to assess accuracy may reveal acuity dependent bias which is not evident from the mean difference alone. If a relationship between accuracy and underlying acuity is suspected, this can be investigated further using formal statistical methods such as regression analysis (see section 4.1.6).

Figure 4.1.3. Bland-Altman plot showing accuracy to be dependent on underlying acuity.



4.1.2. *Precision*

Precision is the repeatability of a test or, in statistical terms, the inverse of the variance of a measurement or estimate. A similar statistical approach to that used to assess accuracy (section 4.1.1) can be used to estimate precision. In this case the pair of measurements taken on each subject are both taken using the same test (or perhaps two equivalent versions of the same test). The difference between test and retest is again calculated by subtracting one measurement from the other. In this case, barring the presence of a 'learning effect' (see section 4.2.7), the mean of the differences would be expected to be approximately zero as both measurements are taken using the same test. For the same reason as stated for accuracy in section 4.1.1, the differences would be expected to conform to a normal distribution. It is expected that 95% of the differences between test and retest will not exceed ±1.96 standard deviations of the distribution of differences. This statistic will be referred to hereafter as the 95% test-retest range (95% TRR). In addition to allowing the precision of different tests to be compared, the fact that the statistic retains the original units is useful because it means that the importance of a measured change in acuity can be interpreted by comparing it with the width of the 95% TRR. From the way in which the 95% TRR is derived, in the absence of

change, we will expect 95% of measured differences between consecutive acuity scores to fall within the 95% TRR. If a measured change in acuity exceeds the 95% TRR we can say (with 95% confidence) that a true change in clinical status has occurred. Therefore, the wider the 95% TRR the larger a measured change must be before it can be attributed to a true clinical change rather than just measurement error.

The Bland-Altman plot can also be used to provide a graphical representation of precision (Fig 4.1.4). The Y-axis this time indicates the magnitude of the difference between test and retest. The points in Fig 4.1.4 are symmetrically distributed about the horizontal midline as would be expected for two measurements taken on the same chart. The vertical spread of points is small indicating small differences between test and retest and therefore a high degree of precision. By contrast, the vertical spread of the points in Fig 4.1.5 is considerably greater suggesting a lower degree of precision. The Bland-Altman plot can also be used to assess whether the degree of precision varies with underlying acuity. Fig 4.1.6 shows a much wider vertical spread of points towards the right side of the plot as compared with the left side. This suggests a reduced degree of precision for those with poorer acuities than for good acuities.

Figure 4.1.4. Bland-Altman plot showing a high degree of precision

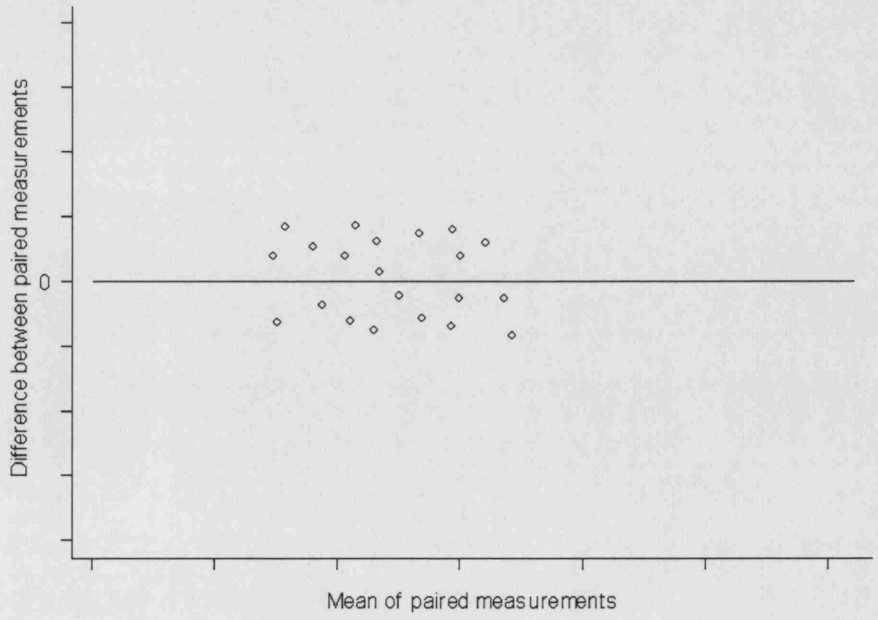


Figure 4.1.5. Bland-Altman plot showing a low degree of precision

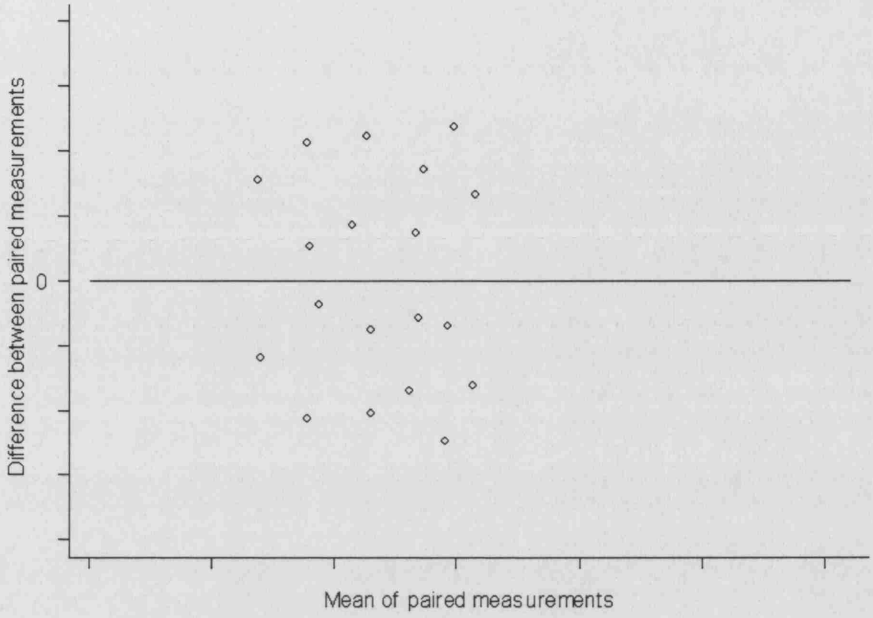
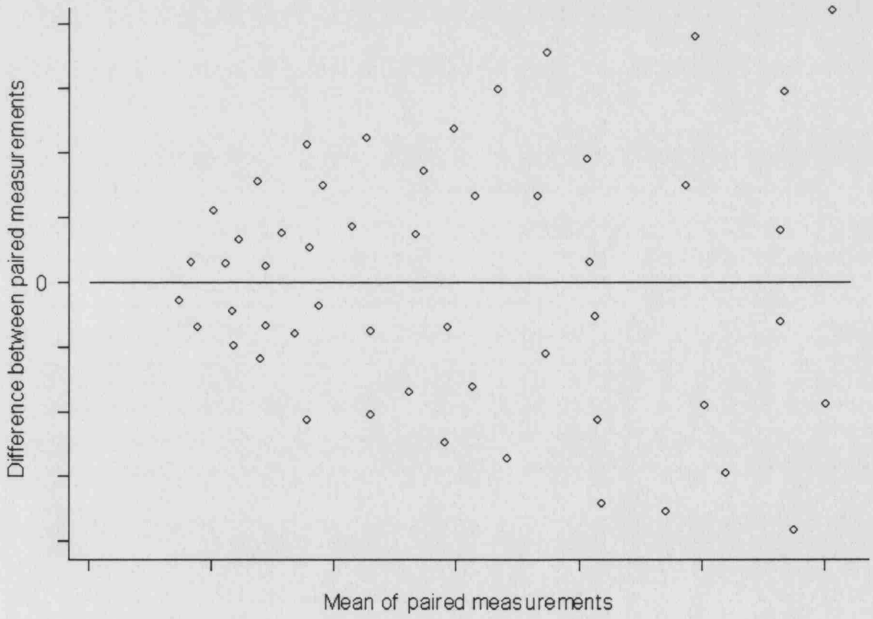


Figure 4.1.6. Bland-Altman plot showing precision to be dependent on underlying acuity



4.1.3. *Agreement*

When a new test is being considered as a potential replacement for another it is the agreement of the two tests which determines the degree to which the tests may be used interchangeably. The degree to which the results of one test agree with those of another depends upon both the accuracy of one test compared with the other, and the relative precision of each test. This fact is illustrated by Fig 4.1.7 and Fig 4.1.8. Fig 4.1.7 is a Bland-Altman plot in which the differences between the acuity scores taken using two different tests are plotted against their mean. The fact that the points are evenly distributed above and below the horizontal midline suggests that the tests are accurate compared with one another (see section 4.1.1). However, the large vertical spread of points indicates a lack of agreement which is due to the relative imprecision of one or both tests. This lack of agreement will limit the ease with which the acuity scores measured using the two tests can be used interchangeably. Fig 4.1.8 depicts the same type of plot as that shown in Fig 4.1.7, but for two different tests. In this case, although the points are again evenly distributed above and below the horizontal midline, the reduced vertical spread of points indicates superior agreement to that seen in Fig 4.1.7. This suggests that the precision of both tests is relatively high.

Figure 4.1.7. Bland-Altman plot: poor agreement regardless of a high degree of accuracy

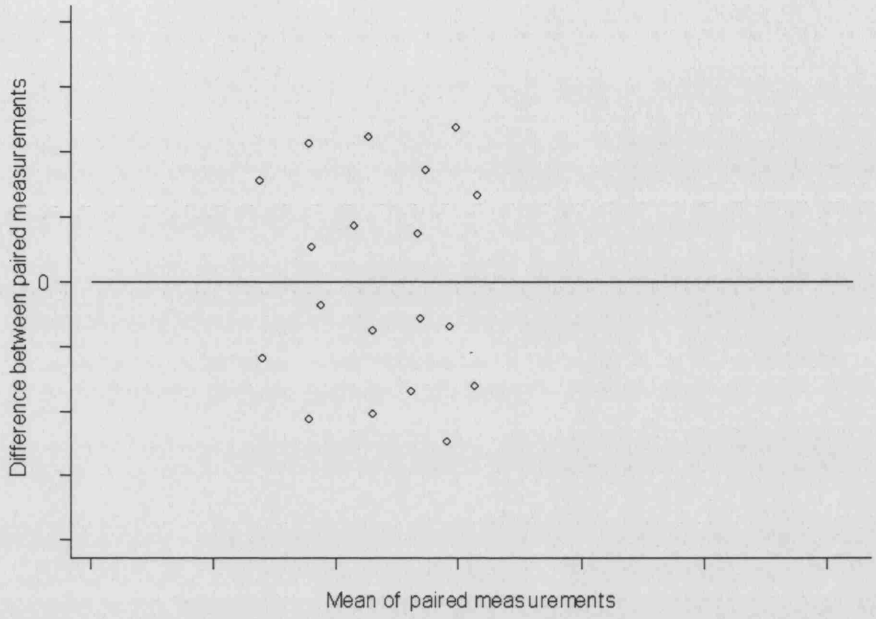
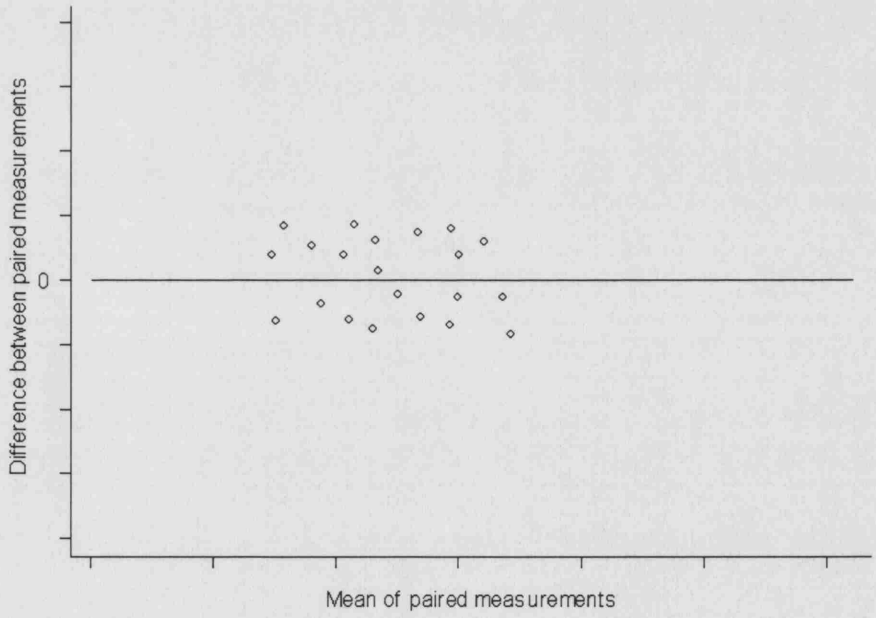


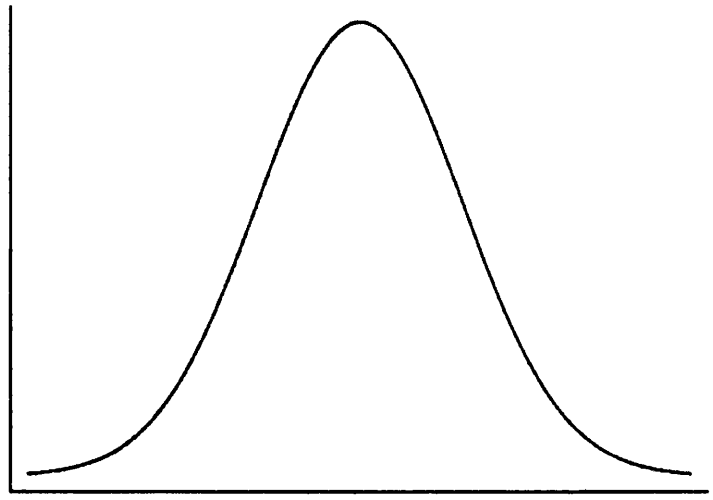
Figure 4.1.8. Bland-Altman plot: good agreement along with a high degree of accuracy



4.1.4. *Prerequisites for the use of the methods of Bland & Altman*

The form of analysis proposed by Bland and Altman uses the distribution of differences between paired measurements. This method assumes the distribution of these differences to be approximately normal. The distribution of differences will often conform to a normal distribution even where the measurements themselves do not, as the distribution is influenced largely by measurement error[202]. In addition, the differences between the paired data should not be related to their mean. In certain circumstances (e.g. where the difference is proportional to the mean) a relationship between the two may be overcome by transforming the data into a form where the relationship is no longer evident. The analysis can then be carried out as normal, and the resultant statistics transformed back into the original units. This will result in confidence limits which are not symmetrical about the mean, but allows a meaningful interpretation of the data.

Figure 4.1.9. The Normal distribution

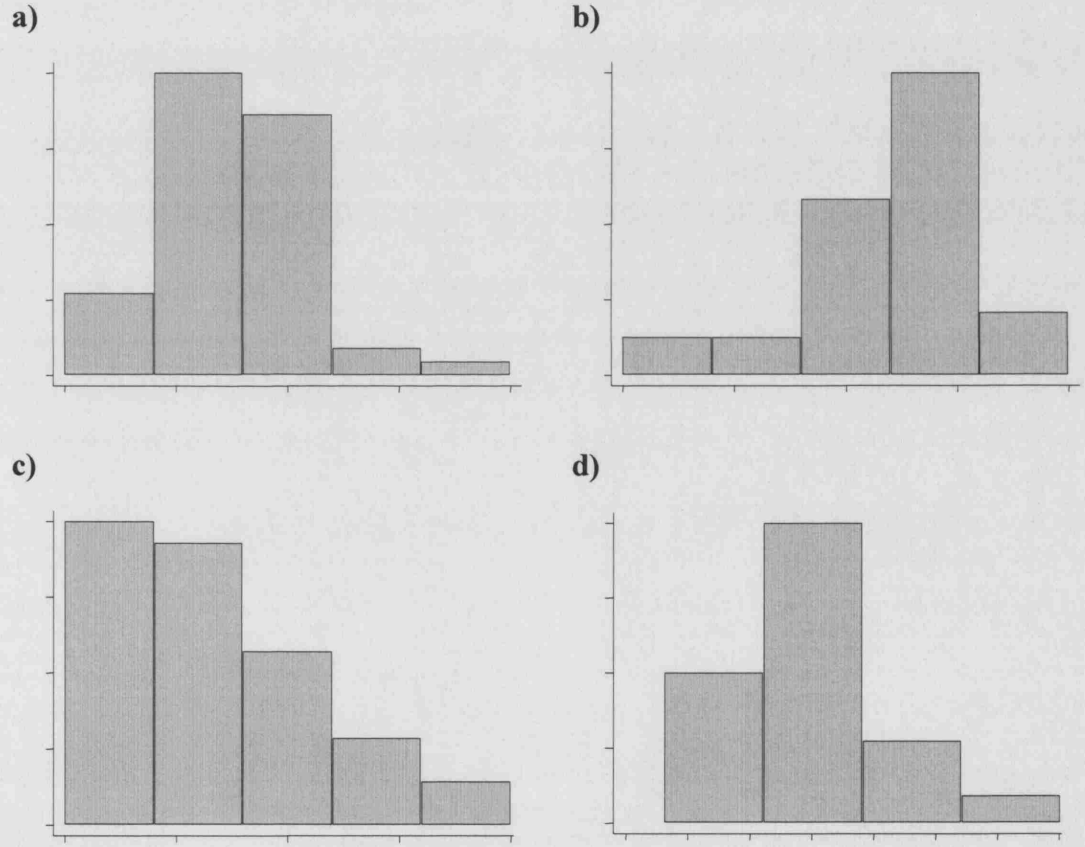


4.1.5. *Assessing normality*

As referred to in section 4.1.4, the validity of the analyses employed in this thesis is dependent upon the data being normally distributed. Parametric statistical methods (those which are based upon a distribution, the shape of which is fully described by one or more parameters) are robust. In other words they are relatively immune to departures from the shape of the distribution in question. It is still however

appropriate that the data be assessed to see how well they conform to, in this case, the normal distribution. Some idea of how the data is distributed can be gained by simply plotting the data in histogram form. The histograms shown in Fig 4.1.10 a and b relate to different data sets but both appear to approximate a normal distribution. Both histograms appear somewhat asymmetric, Fig 4.1.10a having a longer right hand tail, and Fig 4.1.10b having a longer left hand tail. The former is described as displaying a positive skew and the latter a negative skew.

Figure 4.1.10. Using histograms to assess normality

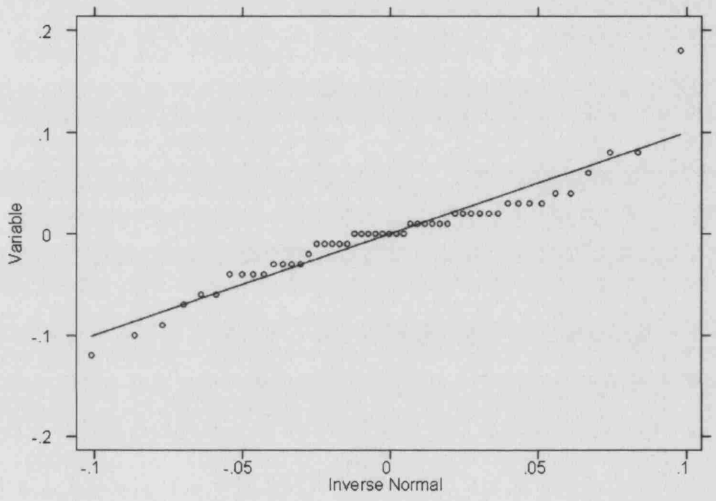


Apart from the subjective nature of judging normality through the inspection of histograms, a further disadvantage is that the appearance of a histogram can be influenced by how the data are grouped into bins (the specified ranges into which the data points are grouped). Figs 4.1.10 c and d are displaying exactly the same data but the use of fewer bins in the latter has created an impression of normality. A better way to graphically assess normality is to use quantile-normal plots in which the variable in

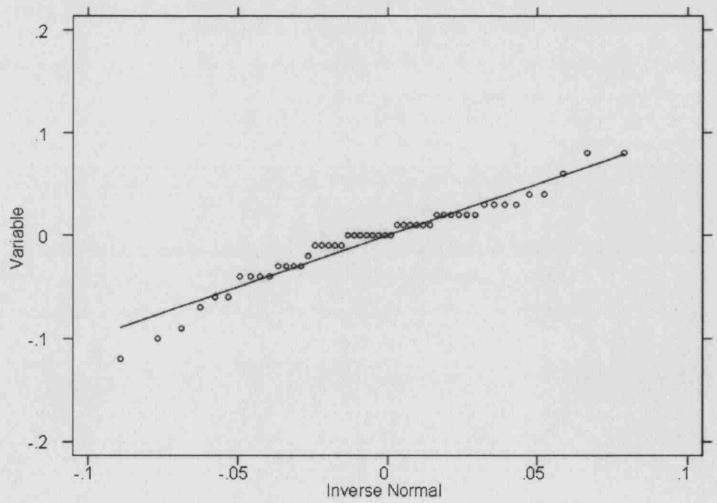
question plotted against the quantiles of the normal distribution. This allows a more reliable assessment of normality as the normal distribution is shown as a straight line. Figs 4.1.11 a and b shows quantile-normal plots generated for the same data as the histograms in Figs 4.1.10 a and b. A large outlier can be seen in the upper right hand corner of Fig 4.1.11 a which is not obvious from the corresponding histogram in Fig 4.1.10 a.

Figure 4.1.11. Using the quantile-normal plot to assess normality

a)



b)



A formal statistical test which seeks evidence of departure from normality is the Shapiro-Wilk test[204]. This test calculates a statistic (W) with a corresponding p-value (see section 4.1.6). The smaller the p-value, the stronger the evidence of a departure from a normal distribution. One advantage of the Shapiro-Wilk test is that it provides an objective and quantitative approach to assessing normality. However, it should be noted that the Shapiro-Wilk test is highly sensitive to departures from normality such that even a single outlier may have a strong effect on the p-value[205]. This can be seen in the case of Fig 4.1.11 b which depicts the same data as 4.1.11 a except the outlier referred to earlier has been removed. The respective p-values for the Shapiro-Wilk test for Figs 4.1.11 a and b are p=0.006 and p=0.158 illustrating that the single outlier was sufficient to result in strong evidence of departure from normality. This thesis utilises the Shapiro-Wilk test to screen for departures from normality. For p-values equal to or greater than 0.05 the data were considered to be normally distributed and no further examination of the distribution was performed. For p-values of less than 0.05, quantile normal plots were used to further examine the distribution and seek outlying data points. In the event of any such outliers, any subsequent Bland-Altman analysis can be conducted with and without these outliers such that their influence upon the results can be determined.

4.1.6. *Other statistical methods utilised in this thesis*

The methods described in sections 4.1.1 and 4.1.2 utilise a statistical approach known as estimation in which conclusions are drawn about all possible subjects (the population) from observations made on a smaller group of individuals (a sample) who are deemed representative of the population. Many of the other statistical methods employed in this thesis utilise an alternative approach known as hypothesis testing, and are therefore often referred to as statistical 'tests' (the Shapiro-Wilk described in section 4.1.5 is an example of a hypothesis test). Hypothesis testing begins with establishing a hypothesis that the effect of interest (e.g. the difference between the means of two sets of observations) is zero. This is known as the 'null hypothesis' as opposed to the 'alternative hypothesis' which states that the effect is not zero. The next step is to calculate the probability that the observed finding could have occurred if the

null hypothesis were true. This is done via the calculation of a 'test statistic' which allows the observed finding to be compared with the known distribution of what would be expected if the null hypothesis were true. The smaller the resultant probability (otherwise known as the 'p-value') the less tenable the null hypothesis. Convention generally dictates that for p-values of less than 0.05, the null hypothesis can be rejected in favour of the alternative hypothesis. In this case, the result is referred to a 'statistically significant'. It should be noted that a p-value in excess of 0.05 does not necessarily mean that the null hypothesis is true, as such a result may also mean that an effect was present, but the study was too small to detect it.

Paired t-test

This is a method of comparing groups of paired observations. Paired data most commonly arise when a group of subjects is studied more than once. For example, the same group of subjects may undergo two blood pressure measurements, separated by one week. In attempting to establish whether blood pressure has changed in the group during the intervening week it is possible to take advantage of the fact that the data are paired. The average difference between observations for each individual, along with the variability of these differences may be used to determine whether blood pressure has changed and inter-subject differences can be ignored. The paired t-test can be used to test the null hypothesis that the mean blood pressure for the group has not changed during that time. Rather than using the normal distribution which is valid only for larger groups of observations, the t-test uses the similar t-distribution which is valid for any sample size.

The F-test ('Fisher test' or 'variance ratio test')

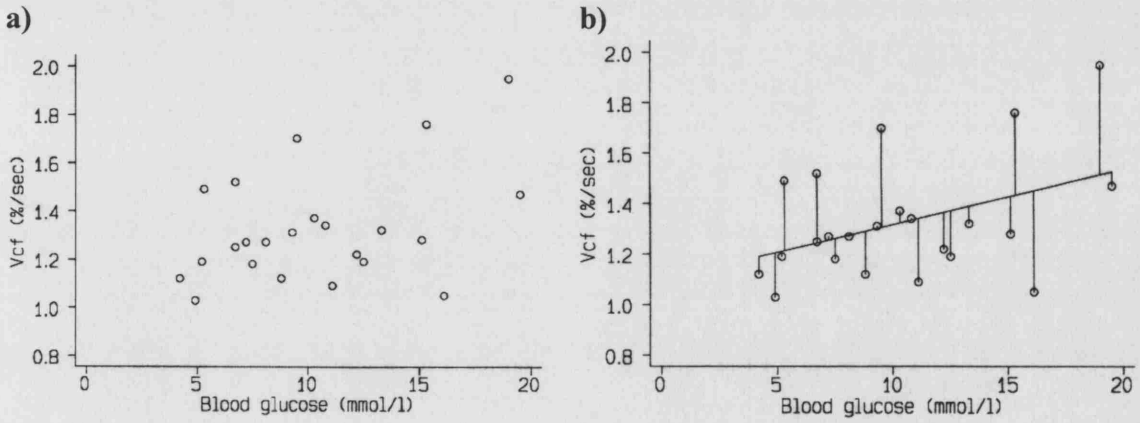
This test performs a similar function to the t-test, but rather than seeking evidence for a difference between the means of two groups of observations, the F-test seeks evidence of a difference in variance. The Bland-Altman method of assessing precision requires that the standard deviation of the distribution of differences between test and retest be calculated (section 4.1.2). If precision is estimated for more than one test, the F-test may be used compare the variance of the distribution (and hence precision) for one test with that of the other. In this case the null hypothesis is that variance of two series of

measurements is the same. The test statistic F is calculated with its corresponding p-value. The smaller the p-value the larger the probability that one test is more precise than the other. For a comparison across more than two groups, Levene's test may be used, which is particularly robust to departures from normality. In this case the null hypothesis is that the variance is the same across multiple groups.

Linear-regression analysis

Regression is a method of describing the relationship between two variables such that one variable (the 'predictor' variable) can be used to predict the other (the 'response' variable). A scatter plot of the two variables is created (with the response variable on the Y- axis and the predictor variable on the X-axis) and a straight line fitted to the data which gives the best prediction of Y for any value of X. Fig 4.1.12 shows the standard method by which this is achieved. Fig 4.1.12a shows the relationship between blood glucose and the mean velocity of circumferential shortening of the left ventricle (VcF). In Fig 4.1.12b the data have been fitted with a straight line such that the sum of the squares of the vertical distances between each point and the line is at its minimum value.

Figure 4.1.12. Fitting two variables with a 'least squares' regression line



This line is referred to as a 'least squares' regression line, and the general equation for a such a regression line is:

$$Y = a + bX$$

where a is the intercept of the line with the Y-axis, and b is the slope of the line. The slope of the line in Fig 4.1.12b suggests that increasing blood glucose is associated with an increase in VcF. Formal statistical methods can be used to test the null hypothesis that the slope of the line is zero. For the data shown in Fig 4.1.12 the test statistic is t = 2.10 and P = 0.05. Hence the relationship between blood glucose and VcF borders on statistical significance.

Sensitivity, specificity and the ROC curve

The diagnostic performance of a test may be described in terms of sensitivity (the proportion of diseased individuals who are correctly identified as such by the test) and specificity (the proportion of normal individuals who are correctly identified as such by the test). Where a single test is used to determine disease status, the test result is typically a continuous variable whose magnitude is associated with the risk of disease. A cut-off is therefore required to categorise subjects as 'diseased' or 'normal'. If there is no overlap between the distribution of the variable in the diseased population and that in the normal population (Fig 4.1.13a), then it is possible to establish a cut-off which will produce both sensitivity and specificity of 100% i.e. perfect diagnostic performance. In practice, this is seldom the case, and a degree of overlap will exist (Fig 4.1.13b).

Figure 4.1.13a. Diseased & normal: non-overlapping distributions

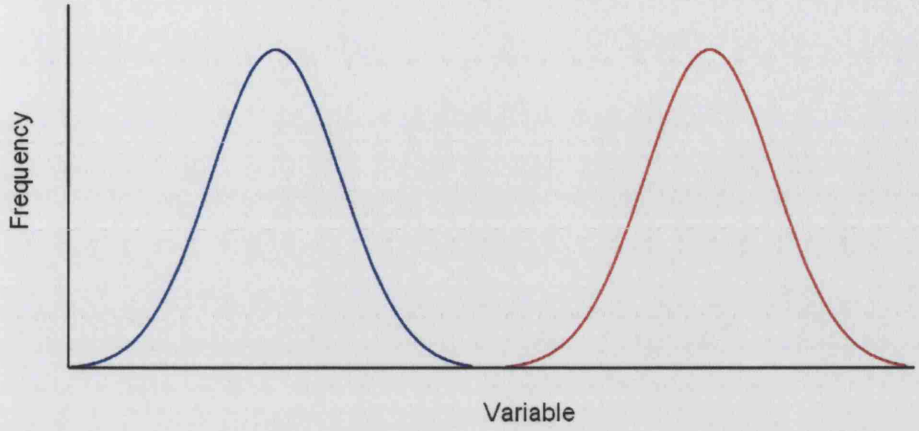
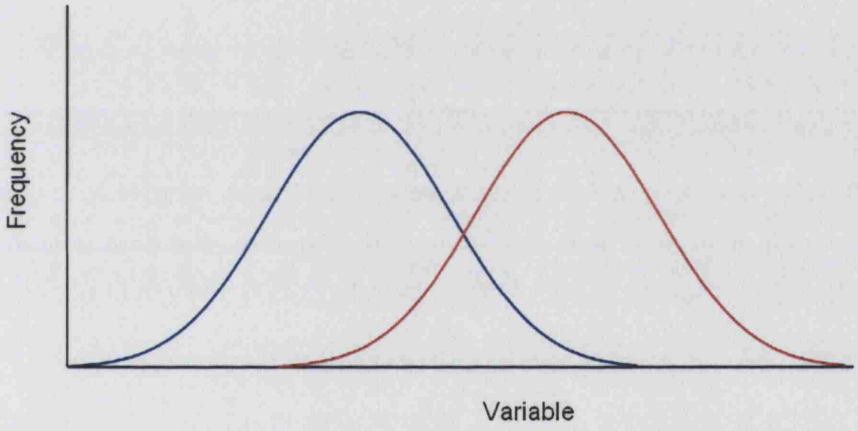


Figure 4.1.13b. Diseased & normal: overlapping distributions

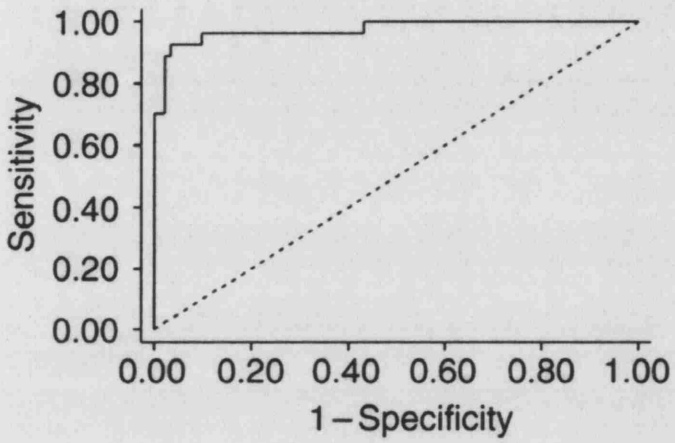


— distribution of the variable in the normal population
— distribution of the variable in the diseased population

In this case, varying the cut-off will influence both sensitivity and specificity such that one is traded off against the other. An overall impression of the diagnostic performance of a test can be gained by plotting sensitivity against specificity (or to be precise 1-specificity) over a range of potential cut-offs. The resultant curve is known as a Receiver Operator Characteristic or ROC curve. A sample ROC curve is shown in Fig 4.1.14. The oblique diagonal line represents the diagnostic performance which could be expected through chance alone (i.e. flipping a coin to decide who is normal and who is diseased). The better the diagnostic performance of the test, the further the curve will move towards the upper left hand corner of the plot. As the curve approaches the upper left corner of the graph the area under the curve will increase,

reaching a maximum of unity for perfect diagnostic performance. The area under the curve can therefore be used to compare the diagnostic performance of one or more tests, and any difference in area can be tested statistically.

Figure 4.1.14. Sample ROC curve



4.1.7. *Software used for statistical analysis*

Basic calculations including the derivation of descriptive statistics such as means, medians, and their confidence intervals were performed using the 'Data Analysis Tool' in Microsoft Excel 2002 (Microsoft Corporation, www.microsoft.com). More complicated statistical analyses such as the Shapiro-Wilk, the F-test and Levene's test were carried out using Stata version 7.0 (Stata Corporation, www.stata.com).

4.2. DATA COLLECTION

4.2.1. *Subjects*

Various recruitment criteria were employed during this research. In all cases, subjects were aged 14 years or over and were deemed capable of giving informed consent and comprehending the requirements of the study protocol.

4.2.2. *Informed consent and ethical approval*

All experiments described in this thesis were conducted in accordance with the tenets of the Declaration of Helsinki[A]. Local ethical approval was given, and informed consent obtained in all cases.

4.2.3. *Number of examiners*

All the visual acuity data presented in this thesis were collected by the author; an optometrist experienced in the measurement of visual acuity.

4.2.4. *Room conditions*

All data collection was carried out within a clinic or office environment lit solely by fluorescent strip lighting, with no external windows. More than one testing environment was utilised during the experiments described in this thesis. However, the test environment and room lighting was kept constant for the duration of each individual experiment. Any equipment utilised in the collection of data was used for the duration of study in question.

4.2.5. *Luminance and contrast*

Details of visual acuity test design are listed separately for each experiment. Although background luminance and contrast varied between experiments, both were kept constant for the duration of each experiment. Chart luminance was measured in Candelas per metre-squared (cd/m^2) using the CS100 luminance meter (Minolta Co Ltd, Japan). All tests were required to achieve a background luminance of at least 80

[A] As adopted by the World Medical Association General Assembly, Helsinki, Finland 1964 and amended by subsequent General Assemblies up to and including October 2000 (www.wma.net/e/policy/b3.htm)

cd/m^2 and a Weber contrast of at least 80%. Weber contrast was calculated using the following formula:

$$\text{Contrast (\%)} = 100\ \{(L_{MAX} - L_{MIN}) / L_{MAX}\}$$

where L_{MAX} is the background luminance and L_{MIN} is stimulus luminance.

4.2.6. *Refractive correction*

One of two forms of refractive correction was used throughout this series of experiments:

Habitual correction: The subject's own current distance spectacle correction established via focimetry of their current distance or multifocal spectacles. This prescription was reproduced in an Oculus Universal trial frame (Keeler Ltd Windsor, UK) using the minimum number of trial lenses with a standard black occluder before the non-tested eye. For subjects who did not possess distance or multifocal spectacles, the trial frame was used solely to provide occlusion to the non-tested eye.

Full refractive correction: The subject underwent formal subjective refraction whilst viewing a visual acuity chart from a distance of 6 metres. The orientation and magnitude of any cylindrical component was established using Jackson crossed cylinders. The spherical component was then refined either through binocular balancing (using the Humphriss Intermediate Contrast technique[206]) or monocularly in the presence of any contraindication to the use of binocular balancing (e.g. amblyopia or asymmetric acuities). Both sphere and cylinder magnitude were established to ±0.25D using an endpoint of maximum plus/minimum minus for maximum visual acuity. All refractions were carried out using a visual acuity chart which was not subsequently used for the purposes of data collection.

4.2.7. *Use of one eye per subject*

For many of the investigations described herein, a high proportion of subjects had two eyes which meet the study inclusion criteria. However, the use of both eyes of each subject can complicate analysis of the data[207]. Unless statistical analyses which take

account of the correlation between two eyes of the same subject are used, the precision of the study, and the statistical significance of any result, will be overestimated. Alternatively, only one eye of each subject may be recruited into the study. This avoids statistical bias without complicating the analysis of the data. Accordingly, all the investigations described in this thesis use only one eye of each subject.

4.2.8. *The ordering of measurements*

It has been shown that on average, a slight improvement in the acuity score occurs when VA measurements are repeated on a given subject. This tendency is referred to as a 'learning effect' or 'memory effect'. Ferris et al[94] demonstrated an average improvement between test and retest of ½ an ETDRS letter (where the retest was within 24 hours of the initial test). This is consistent with the results of other studies which have shown a learning effect to be present, but small in magnitude[74 81 208]. McMonnies has shown that learning effects may persist for up to 10 days and are most likely due to familiarity with the letter subset used rather than familiarity with the procedure per se[209].

As all investigations involved multiple measures of acuity, the order of presentation was randomised to remove learning or fatigue effects as a potential source of bias. In addition, no repeated measurements were taken using the same letter combinations unless separated in time by at least two weeks. The only exception to this was where the number of ETDRS measurements required in a single session exceeded three, therefore at least one chart was used twice. In this event, a restriction to the randomised test order was introduced such that the same chart was not used for consecutive measurements. Also, one of the two measurements taken with the same chart was conducted with each line being attempted in reverse.

Randomisation was carried out using series of 'pseudo-random' numbers generated by computer. This form of simple randomisation was thought to be adequate for the purposes of avoiding ordering effects.

4.2.9. *Termination rules and forced choice procedures*

As referred to in section 2.10.7, the use of a rigorous rule which dictates when an acuity measurement is terminated eliminates some potential sources of bias, and results in a more reliable result. Carkeet has suggested that when using logMAR charts based upon the Bailey-Lovie design, the measurement should not be terminated until the subject has misnamed at least 4 of the 5 letters on a line[114]. For the purposes of these studies, a conservative termination rule of a full line of errors was employed. Although slightly more conservative, this rule provides the opportunity of rescoring a given measurement with a less conservative rule, something which cannot be done in reverse. The use of a more conservative rule also allows for the fact that the optimum rule in Carkeet's study may be specific to the particular sample used in his study. The conservative approach therefore allows a margin for error.

Use of the kind of termination rule referred to above requires the subject to attempt a number of sub-threshold stimuli, which are by definition too small to resolve. The reluctance of many subjects to do this necessitates the use of forced choice methods which require the subject to continue attempting stimuli until the termination rule is fulfilled (see section 2.10.7).

4.2.10. *Scoring procedures*

Two scoring methods were employed in these investigations: the interpolated method and the line-assignment method.

Line-assignment scoring: This method was referred to in Section 2.10.4. When using a line-assignment method, the only possible acuity scores are those corresponding to the various stimulus sizes featured on the chart. For the purposes of this thesis, a line-assignment acuity score refers to the value of the smallest line on which at least half of the stimuli are named correctly.

Interpolated (single-letter)scoring: The use of interpolation allows for acuity scores which fall between the score corresponding to the various stimulus sizes, thereby producing a finer measurement scale. This is done by giving credit for each stimulus correctly named (see section 2.10.6). An interpolated logMAR acuity score ('single-letter' score) is produced using the formula:

$$\text{Visual acuity (logMAR)} = +1.10 - T_C L_V$$

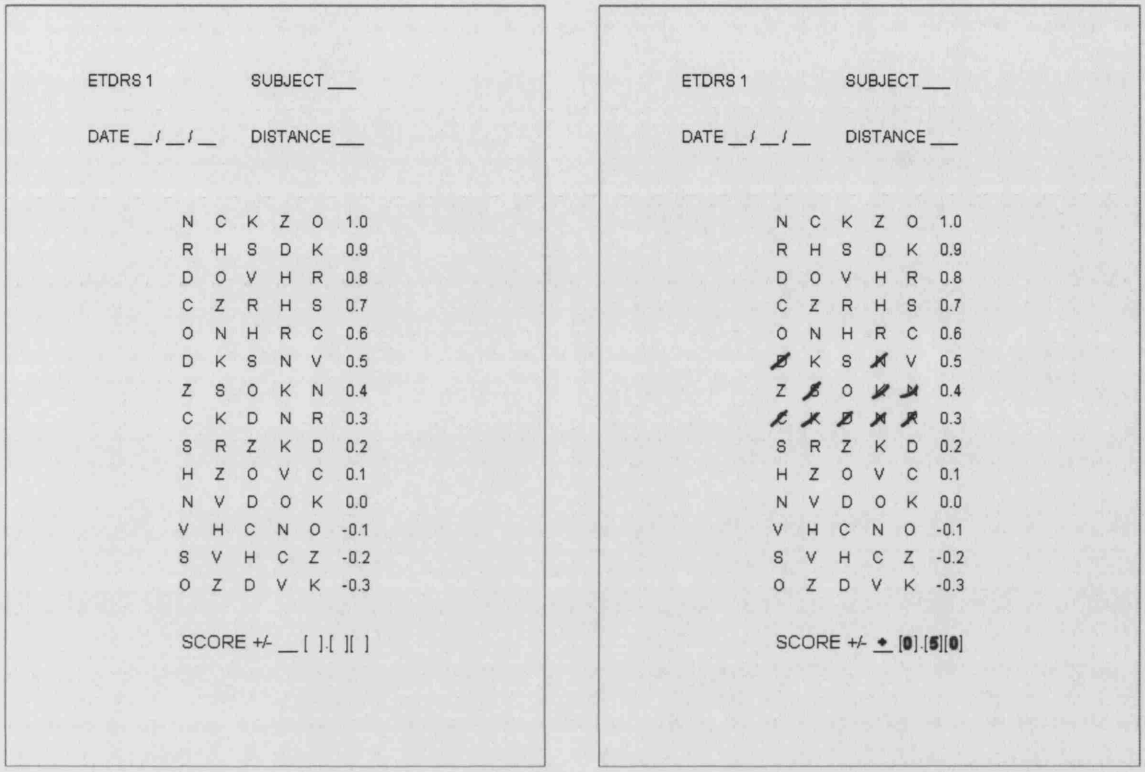
where T_C is the total number of correctly named letters, and L_V is the logMAR value of each letter on the chart. The individual letter values are produced by dividing the logMAR interval between a given line and the line below by the number of letters on that line[94].

4.2.11. *Recording subject responses*

During each 'conventional' (i.e. non-computerised) acuity measurement, the subject's responses were recorded on a 'data proforma' (Fig 4.2.1). Each proforma featured the letters from the relevant visual acuity chart in order grouped in lines as per the chart's design. Also featured is the chart used, the subject's study number, the date and the distance at which the measurement was carried out. Other details relating to the subject such as date of birth and refraction were recorded elsewhere. During a measurement, the examiner followed the subject's responses on the proforma as each optotype was attempted in turn. For each incorrect response, the examiner crossed the appropriate letter through with an oblique line. As a termination rule of a full row of errors was used (section 4.2.8) the measurement was terminated once all the letters on a line were crossed through. The examiner then moves straight on to the next measurement in the series. Recording the subjects responses in this way allowed the score to be calculated following the completion of the data collection session. This record allowed for an acuity to be subsequently recalculated, either to rule out an error, or to assess the effect of a different scoring method on the measurement. The proforma for each measurement was printed on an A4 sheet, and the sheets were stapled together in order as determined by randomisation (see section 4.2.7), thereby minimising the chances of an ordering error on the part of the examiner.

Figure 4.2.1. Sample data proformas for version 1 of the ETDRS chart:

a) Uncompleted **b) Completed**



4.2.12. *Other procedures followed during data collection*

A number of other procedures were followed for all the experiments described herein in an attempt to minimise the influence of extraneous variables. The subject was seated for all acuity measurements, and the position of the subject kept constant for the duration of each study. The subject was refracted, or their spectacles focimetered, according to the requirements of the individual study (see section 4.2.5). A chart was used to explain the nature of the test emphasising the use of forced choice methodology and termination rules. Subjects were also explicitly asked:

- to remain as still as possible, and refrain from leaning, squinting, or tilting of the head,
- to keep both eyes open, even though one is occluded,
- to avoid pausing for long periods over a given letter so as to keep attempting letters at a steady rate, and

- to alert the examiner if they feel they may have 'lost their place' on the chart (in this event, the examiner was allowed to indicate the next stimulus using a black pointer held beneath the letter, but no closer to the stimulus than any surrounding contour).

Care was taken to avoid the subject seeing any of the study charts in advance of the appropriate time.

5. THE DEVELOPMENT OF AN ABBREVIATED logMAR VISUAL ACUITY CHART FOR ROUTINE CLINICAL USE

5.1. BACKGROUND & AIMS

Background

The deficiencies of Snellen's chart design are well known (see section 2.11). Many of the problems with Snellen's design have been addressed through the development of the current gold standard charts based upon the Bailey-Lovie design. The advantages of the design principles employed in logMAR visual acuity tests such as the Bailey-Lovie and ETDRS charts over the Snellen chart have also been well described (see section 2.12). Despite these advantages, uptake of these charts in routine clinical practice has been minimal. There are a number of possible explanations for this. Firstly, clinicians are very familiar with the Snellen form of notation, and may be reluctant to change to a form of notation with which they are unfamiliar, and which, initially at least, is less meaningful to them. It is also possible that contemporary logMAR charts are perceived as being cumbersome to use in comparison with a Snellen chart. Another possible barrier to more widespread uptake of these charts is that clinicians may not be aware of the limitations of the Snellen chart, and may therefore feel that logMAR charts will offer them no advantage in clinical practice. Finally, although the prices of the Snellen and ETDRS tests (including lightbox) are similar, the cost of replacing Snellen charts may be prohibitive, as a switch to Bailey-Lovie type charts requires a different lightbox to be purchased.

It is hypothesised that an abbreviated version of the ETDRS logMAR chart design will allow logMAR measurements to be taken in a clinically acceptable period of time, with a greater degree of precision than the Snellen chart. An abbreviated ETDRS logMAR chart design may therefore allow the benefits of contemporary logMAR chart designs to be utilised in routine clinical practice.

Aims

- To assess the accuracy of three prototype visual acuity charts (featuring abbreviated versions of the Bailey-Lovie design), and the Snellen chart, compared with the ETDRS chart (which acted as the gold standard).
- To compare the precision of the prototype and Snellen charts with that of the gold standard ETDRS chart.

5.2. METHODS

5.2.1. *Subjects*

Subjects were recruited from an outpatient ophthalmic clinic population. The inclusion criteria were as follows:

- Snellen acuity of between 6/5 and 6/60 (as recorded from hospital notes on the day of recruitment),
- a diagnosis of either cataract, pseudophakia or early glaucoma (with no visual field defects within the central 20°), and
- able to understand and comply with the testing protocol.

Only one eye of each subject was used for the study to avoid any statistical bias (see section 4.11). Where both eyes met the inclusion criteria, the eye with the poorest acuity was used as the study eye. This was done to broaden the distribution of acuity in the study. A minimum of 40 subjects were recruited. The choice of 40 subjects was a pragmatic one but is consistent with the size of previous similar studies by other workers.

5.2.2. *The charts*

5 different types of visual acuity chart were employed in the study. Two versions of each of the five different chart designs were used to avoid memorization effects, making a total of 10 charts. The chart design parameters and the acuity range at a 6 metre testing distance is summarised in Table 5.2.1, and the appearance of the charts is depicted in Fig 5.2.1. The number of letters per line, line interval and single letter value are constant across the chart for each chart design except the Snellen chart where the range of values has been given.

The ETDRS chart (Lighthouse International, New York)

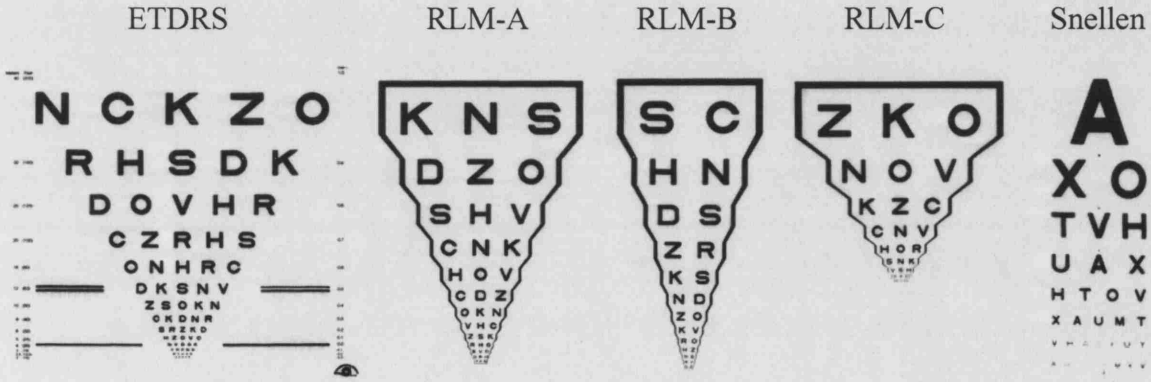
This chart is described in full in Section 2.13. ETDRS charts 1 and 2 were used. The refraction chart was used for patient demonstration purposes. The charts were retro-illuminated using the dedicated display box achieving a luminance of 300 cd/m^2 and a contrast of 98%.

Table 5.2.1. Summary of chart design.

Chart	No of letters per line	Line interval (logMAR)	Scale increment (logMAR)	Total no of letters	Acuity range at 6m (logMAR)
ETDRS	5	0.1	0.02	120	+0.82 to -0.48
RLM-A*	3	0.1	0.033	39	+0.72 to -0.48
RLM-B*	2	0.1	0.05	26	+0.72 to -0.48
RLM-C*	3	0.15	0.05	27	+0.72 to -0.48
Snellen	1 to 8	0.08 to 0.22	0.01 to 0.11	45	+1.00 to -0.18

*prototype chart

Figure 5.2.1. The charts



Reduced logMAR charts (RLM charts A, B and C)

Three prototype Reduced logMAR (or RLM) chart designs were produced, each representing an abbreviation of the ETDRS design. Two versions of each design were created, differing only in the arrangement of letters used. The RLM chart designs followed that of the ETDRS version of the Bailey-Lovie chart wherever possible. The RLM charts featured a geometric progression of sizes, with a constant inter-line size increment. Inter-letter spacing was equivalent to one letter width, and the inter-line spacing, equivalent to the height of the letters in the lower row. As per the ETDRS chart design, Sloan's own data[98] concerning the relative legibility of the letters in her letter set (Table 2.13.1) were used to arrange the letters such that the total difficulty of each line was approximately equal. For example, Table 5.2.2 shows the total difficulty for each row of the RLM-A chart. For each row, the total difficulty was within ±2% of

the average difficulty of the Sloan letter set. Words and familiar acronyms were avoided. Reducing the number of letters per line reduces the proportion of letters which are subject to contour interaction from both sides (see section 2.10.8). Such an effect has the potential to introduce systematic bias when acuities measured with an abbreviated chart are compared with those measured using the ETDRS chart (see section 2.10.8). In an attempt to avoid this as a source of inaccuracy, all the RLM charts featured a continuous 'crowding bar' to compensate for the reduction in the proportion of letters crowded from both sides (see Fig 5.2.1). The position and dimensions of the crowding bar followed the same geometric progression as the letter stimuli such that the bar's position was 2.5 stroke widths (half a letter) from the end of each row of letters, and its thickness was 1 stroke width (Fig 2.5.1). The 2.5 stroke width separation was chosen based upon previous work which has shown the difficulty of letters crowded in this way to equate well to that of letters in a linear arrangement (see section 2.10.8). The charts were retro-illuminated in the same lightbox as the ETDRS chart achieving a luminance of 560 cd/m^2 and a contrast of 98%.

Table 5.2.2. Difficulty of each line of the RLM-A chart relative to the average difficulty of the Sloan letter set.

Line	Relative letter difficulty (%)			Relative line difficulty (%)
1	K (0.1)	N (9.6)	S (-11.4)	-1.7
2	D (-0.5)	Z (12)	O (-11)	+0.5
3	S (-11.4)	H (7.3)	V (2.3)	-1.8
4	C (-10.6)	N (9.6)	K (0.1)	-0.9
5	H (7.3)	O (-11)	V (2.3)	-1.4
6	C (-10.6)	D (-0.5)	Z (12)	+0.9
7	O (-11)	K (0.1)	N (9.6)	-1.3
8	V (2.3)	H (7.3)	C (-10.6)	-1.0
9	Z (12)	S (-11.4)	K (0.1)	+0.7
10	R (4.3)	H (7.3)	O (-11)	+0.6
11	K (0.1)	V (2.3)	D (-0.5)	+1.9
12	S (-11.4)	R (4.3)	H (7.3)	+0.2
13	V (2.3)	N (9.6)	S (-11.4)	+0.5

The Snellen charts (Clement Clarke UK)

The design of commercially available Snellen charts varies considerably (see Fig 2.12.1). Charts may vary in terms of the number of letters per line, the number of lines, the size increment between lines, the spacing between lines, as well as the selection and style of optotypes. For the study, two Snellen charts were selected to be as similar as possible to prevent the introduction of any extraneous variables. Both charts had non-serif letter optotypes, and were equivalent in terms of number of rows of letters, the number of letters per row, and the horizontal and vertical spacing between letters. Although not all Snellen charts include letter sizes smaller than 6/6, the charts chosen for the study included both a 6/5 line and a 6/4 line. This was done because acuities beyond 6/6 are common and the lack of any letter stimuli beyond this size can artificially reduce an estimate of precision by reducing the discrepancy between test

and retest. Both Snellen charts were retro-illuminated in the standard Clement Clarke lightbox achieving a luminance of 790 cd/m^2 and a contrast of 99%.

5.2.3. *Testing protocol*

Timed acuity measurements were taken on one eye of each subject, one measurement being taken using both versions of each of the five chart types. The time was measured from the subject starting to read the letters until they had made a complete line of errors. Calculation of acuity scores was performed separately to avoid influencing the time measurements. The charts were viewed in random order and responses were recorded on a dedicated data proforma (see Fig 4.2.1).

All measurements were conducted under the conditions described in section 4.2.1. All subjects wore their habitual spectacle correction, or no correction in the event of the subject not habitually using distance spectacles. All charts were read from a distance of 6 metres unless the subject misnamed any letters on the top line of a given chart. In this event, the subject was moved to 1.5 metres and the remainder of the measurements taken at this distance. Forced choice procedures and termination rules were used as described in section 4.2.8.

5.2.4. *Scoring*

Prior to score calculation, the Snellen fraction corresponding to each letter size on the Snellen chart (6/4 to 6/60) was converted into the logMAR format. This is done by taking the log to the base 10 of the reciprocal of the decimalised Snellen fraction. This transform can be performed regardless of the denominator or numerator:

$$\text{i.e. } \log_{10}(1/S) \text{ where S is the decimalised Snellen fraction}$$

$$\text{e.g. Snellen } 6/9 \equiv \log_{10}(1/0.667) = +0.18 \text{ logMAR}$$

An interpolated logMAR acuity score was produced for each chart using the method described in section 4.2.9. This was done for the Snellen chart as well as for the logMAR charts. Due to the irregular progression of letter sizes and the varying number of letters per line on the Snellen chart, this process was considerably more time consuming than for the logMAR charts. The Snellen charts were also scored using the

'line-assignment' method as described in section 4.2.9. This is the most common scoring method employed with the Snellen chart in clinical practice. All acuity scores were adjusted for the use of non-standard testing distances prior to analysis.

5.3. RESULTS

41 subjects were recruited, 16 of whom were male and 25 female. Age at last birthday ranged from 49 to 89 years. 30 subjects had cataract, 7 were pseudophakic and 4 had primary open angle glaucoma. Acuity as measured with the ETDRS logMAR chart ranged from +0.82 to -0.14 logMAR (Snellen equivalent 6/40 to 6/4.3) with a median of +0.34 logMAR (Snellen equivalent 6/13).

Table 5.3.1. Accuracy compared with reference standard ETDRS chart

Charts under comparison	Mean difference (logMAR)	95% c.i. (mean) (logMAR)	Slope of regression line
ETDRS (interpolated) – RLM-A (interpolated)	0.00	(-0.03 to +0.03)	-0.05 (t=-0.85, p=0.399)
ETDRS (interpolated) – RLM-B (interpolated)	+0.01	(-0.03 to +0.05)	-0.01 (t=-0.09, p=0.931)
ETDRS (interpolated) – RLM-C (interpolated)	-0.01	(-0.05 to +0.03)	-0.05 (t=-0.71, p=0.482)
ETDRS (interpolated) – Snellen (line-assignment)	+0.07	(+0.02 to +0.12)	-0.19 (t=-2.39, p=0.022)

Accuracy assessed using mean difference plus 95% confidence interval
t – calculated using linear regression analysis

Accuracy

Table 5.3.1. shows the mean of the distribution of differences between paired measurements taken with the ETDRS chart and the chart under comparison. For the RLM-A, RLM-B, and RLM-C charts, the mean difference was within ±0.01 logMAR (½ an ETDRS letter) of zero suggesting that on average acuities measured using these charts are accurate compared with those measured using the reference standard ETDRS chart. In addition, the 95% confidence interval for the mean included zero in each case providing no evidence of a significant degree of bias between any of the RLM charts and the ETDRS chart.

For the Snellen chart, the mean difference was +0.07 logMAR which suggests that, on average, acuities measured using the Snellen chart are 0.07 logMAR (or approximately 2/3rds of a line of ETDRS letters) better than those measured using the reference standard ETDRS chart. In addition, the 95% confidence interval for the mean excluded zero (+0.02 to +0.12 logMAR) which is suggestive of a statistically significant difference.

Acuity-related bias

To assess the data for evidence of differential bias according to the underlying level of acuity, Bland-Altman plots (see section 4.1.1) were created for each chart design as compared with the reference standard ETDRS chart (Figs 5.3.1-4). These plots show the difference between acuities measured using the two tests plotted against their mean. A visual inspection of the plots for the RLM charts (Figs 5.3.1-3) gave no indication of a systematic variation in accuracy with underlying acuity for any of the prototype charts. A linear regression analysis (see section 4.1.6) of the difference between the acuities measured with the two tests against their mean gave no evidence of a statistically significant relationship between the difference and the mean for any of the RLM charts (see Table 5.3.1).

For the Snellen chart, inspection of the Bland-Altman plot (Fig 5.3.4) suggested a relationship between the difference and the mean such that Snellen acuity scores tended to be more negative (better) than ETDRS acuities for good acuities but similar for poor acuities. This trend can be seen in Fig 5.3.4 as a tendency for points toward the left hand side of the plot to be situated above the midline, whereas those to the

right are more evenly distributed. Linear regression analysis showed this trend to be statistically significant at the 5% level ($t=-2.39$, $p=0.022$).

Figure 5.3.1.Bland-Altman plot for ETDRS (interpolated) vs RLM-A (interpolated)

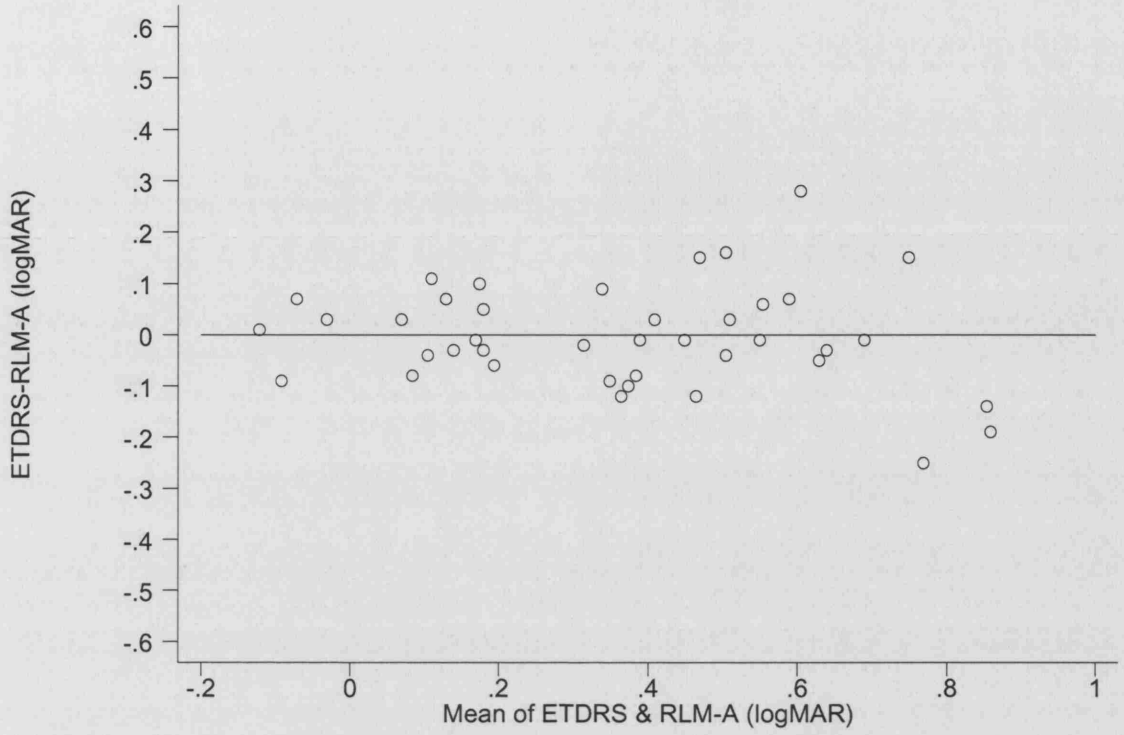


Figure 5.3.2. Bland-Altman plot for ETDRS (interpolated) vs RLM-B (interpolated)

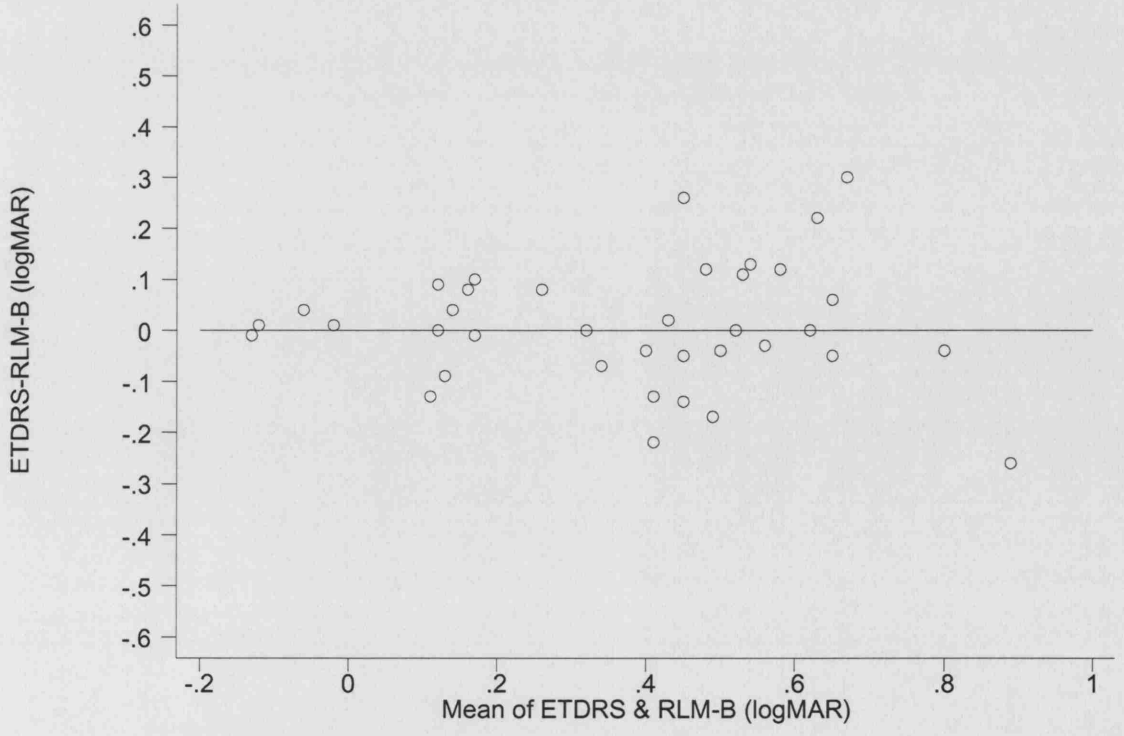


Figure 5.3.3. Bland-Altman plot for ETDRS (interpolated) vs RLM-C (interpolated)

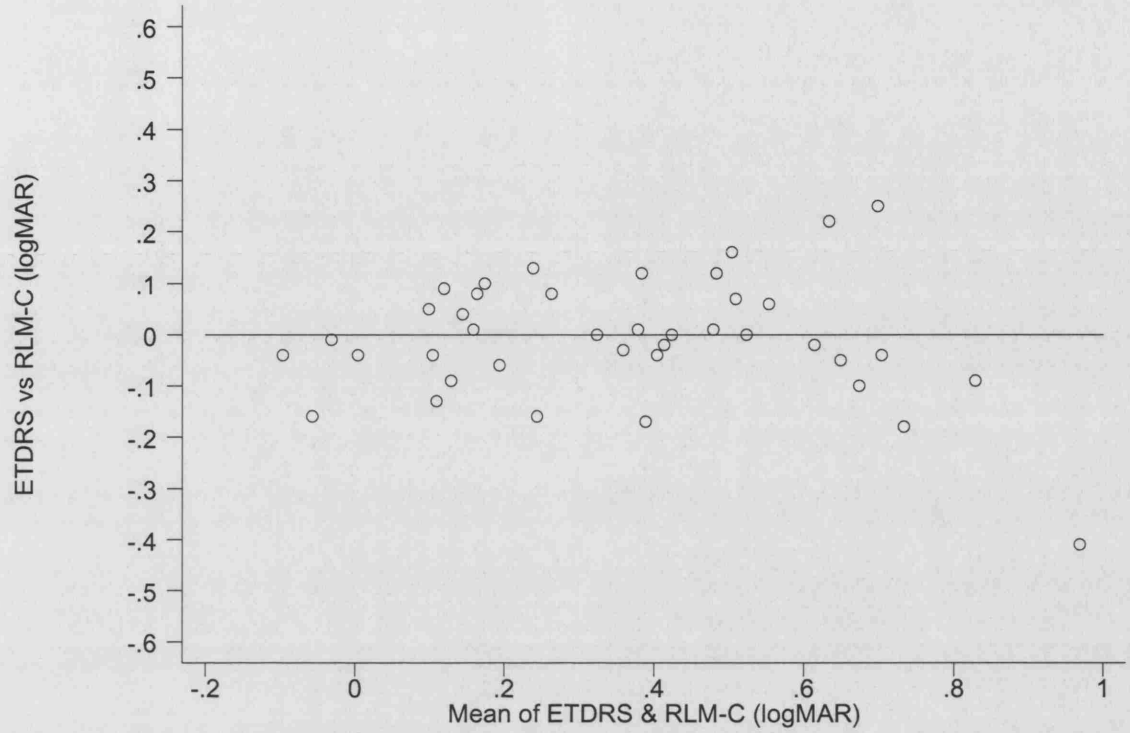
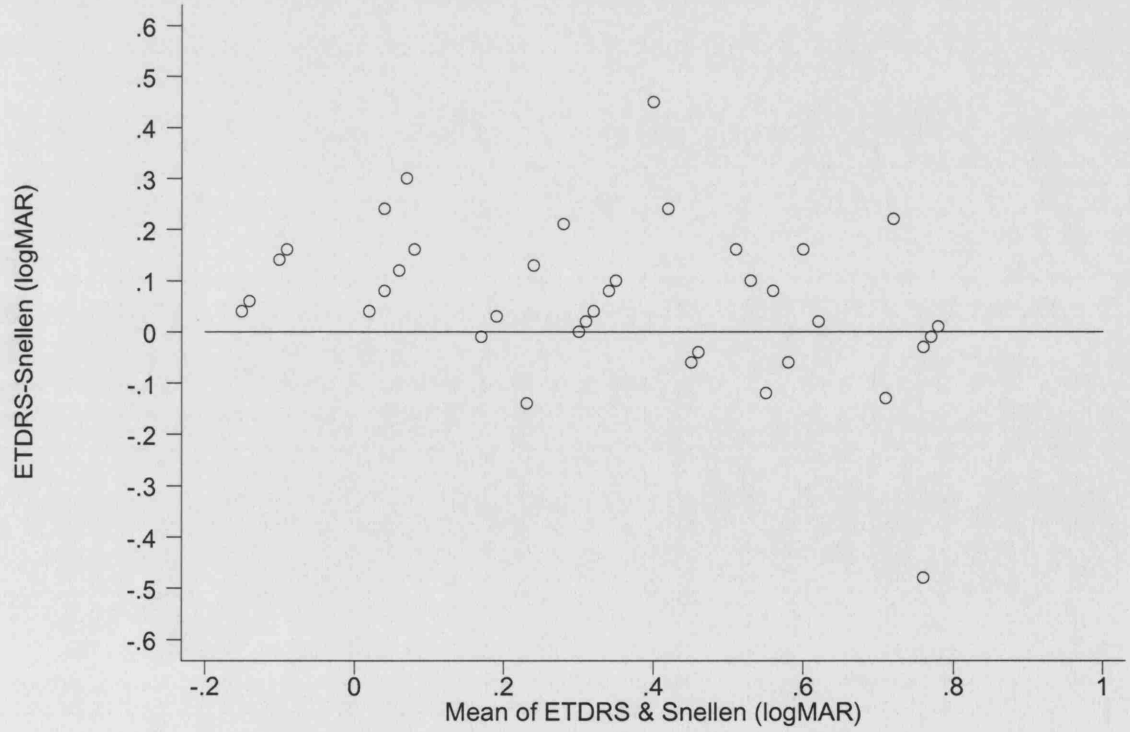


Figure 5.3.4. Bland-Altman plot for ETDRS (interpolated) vs Snellen (line-assign')



The effect of scoring method

The accuracy of the Snellen chart as compared with the gold standard ETDRS chart was assessed using different scoring methods for each chart. Scoring the ETDRS chart with the recommended interpolated method, and the Snellen chart with the typical line-assignment method allows a meaningful comparison of the two charts as they would normally be used. However, it is feasible that the different scoring method may have been at least partly responsible for the observed degree of bias between Snellen and ETDRS acuities. To assess the influence of scoring method, the accuracy of the Snellen chart compared with the ETDRS chart was reassessed using the same scoring method for both charts. This required the Snellen acuity scores to be recalculated using the interpolated scoring method, and the ETDRS scores to be re-calculated using the line-assignment scoring method. Table 5.3.2 shows that accuracy improved slightly when both charts were scored using the line-assignment method, and improved further when both were scored using the interpolated method[A].

Table 5.3.2. The effect of scoring method upon accuracy

Charts under comparison	Mean difference (logMAR)	95% c.i. (mean) (logMAR)	Slope of regression line
ETDRS (interpolated) – Snellen (line-assignment)	+0.07	(+0.02 to +0.12)	-0.19 (t=-2.39, p=0.022)
ETDRS (interpolated) – Snellen (interpolated)	+0.02	(-0.02 to +0.06)	-0.16 (t=-1.95, p=0.059)
ETDRS (line-assignment) – Snellen (line-assignment)	+0.05	(0.00 to +0.10)	-0.21 (t=-2.22, p=0.032)

The results shown in Table 5.3.2 suggest that at least one chart produces acuity scores which are dependant upon the scoring method. To investigate this further, the difference between line-assignment and interpolated acuity scores was calculated (along with its 95% confidence interval) for each chart. Table 5.3.3 shows that the mean difference between line-assignment and interpolated acuities for the ETDRS

[A] Linear regression analysis was suggestive of a reduced degree of acuity-related bias when the two charts were scored using the same method – see the final column of Table 5.3.2.

chart is small (-0.01 logMAR), and the 95% confidence interval for the mean includes zero. This suggests that ETDRS acuities are independent of scoring method. For the Snellen scores, the mean difference is larger (-0.04 logMAR) and its 95% confidence interval excludes zero. This suggests that for the Snellen chart, the line-assignment scoring method produces better acuity scores that the interpolated method.

Table 5.3.3. The effect of scoring method upon ETDRS and Snellen scores

Chart	Mean difference logMAR)	95% c.i. (mean) (logMAR)
ETDRS (line-assignment – interpolated)	-0.011	-0.030 to +0.008
Snellen (line-assignment – interpolated)	-0.041	-0.066 to -0.016

Precision

To assess the suitability of the methods of Bland and Altman for estimating test precision (see section 4.1.2), the distributions of differences between paired measurements were assessed for evidence of departure from a normal distribution using the Shapiro-Wilk W-test (see section 4.1.2). The results of this analysis are shown in Table 5.3.4.

Table 5.3.4. Assessing test-retest data for non-normality

Chart	Scoring method	W
ETDRS	Interpolated	0.996 (p=0.999)
RLM-A	Interpolated	0.949 (p=0.065)
RLM-B	Interpolated	0.963 (p=0.193)
RLM-C	Interpolated	0.868 (p<0.001)
Snellen	Line-assignment	0.941 (p=0.033)

This analysis suggested weak evidence (p=0.033) against the hypothesis of normality for the distribution of differences between test and retest for the Snellen line-assignment measurements, and stronger evidence (p<0.001) against normality for the equivalent distribution for the RLM-C chart. Quantile-normal plots (see section 4.1.5) were generated for these distributions to enable a more detailed evaluation (see Figs 5.3.5 and 5.3.6). For both the RLM-C and Snellen charts, the departure from normality was attributable in the main to a single outlying point (highlighted by a red arrow in Fig 5.3.5 and Fig 5.3.6). For the RLM-C chart, the outlying point corresponded to subject 28, whereas for the Snellen chart, it corresponded to subject 2. Repeating the Shapiro-Wilk test with these points excluded gave no evidence of departure from a normal distribution for either the RLM-C chart (W=0.963, p=0.19), or the Snellen chart (W=0.977, p=0.56) at the 5% level.

Figure 5.3.5. Quantile-normal plot for the distribution of test-retest differences for the RLM-C chart

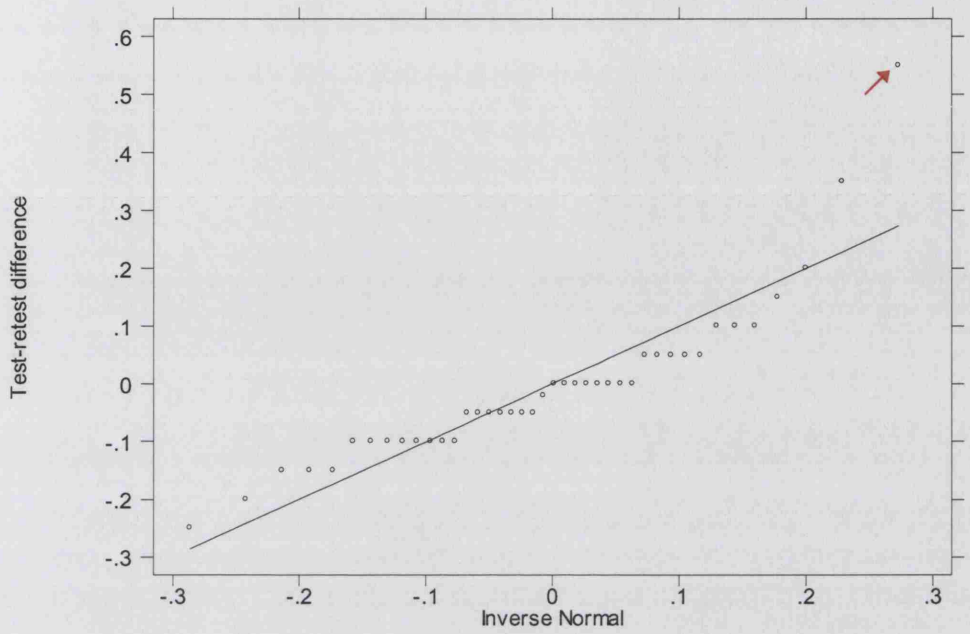
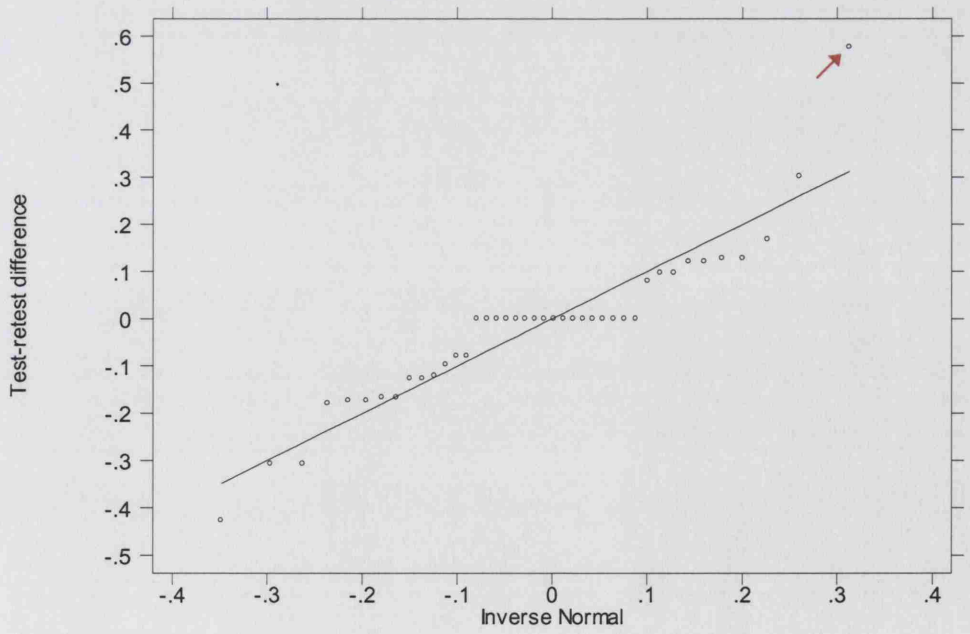


Figure 5.3.6. Quantile-normal plot for the distribution of test-retest differences for the Snellen chart



Bland-Altman plots for precision were generated from the test-retest data for all 41 subjects for each chart (Figs 5.3.7-12.). In general, the vertical spread of test-retest points was least for the ETDRS chart, and greatest for the Snellen chart, with the RLM charts lying in between.

Figure 5.3.7. Bland-Altman plot for precision – ETDRS

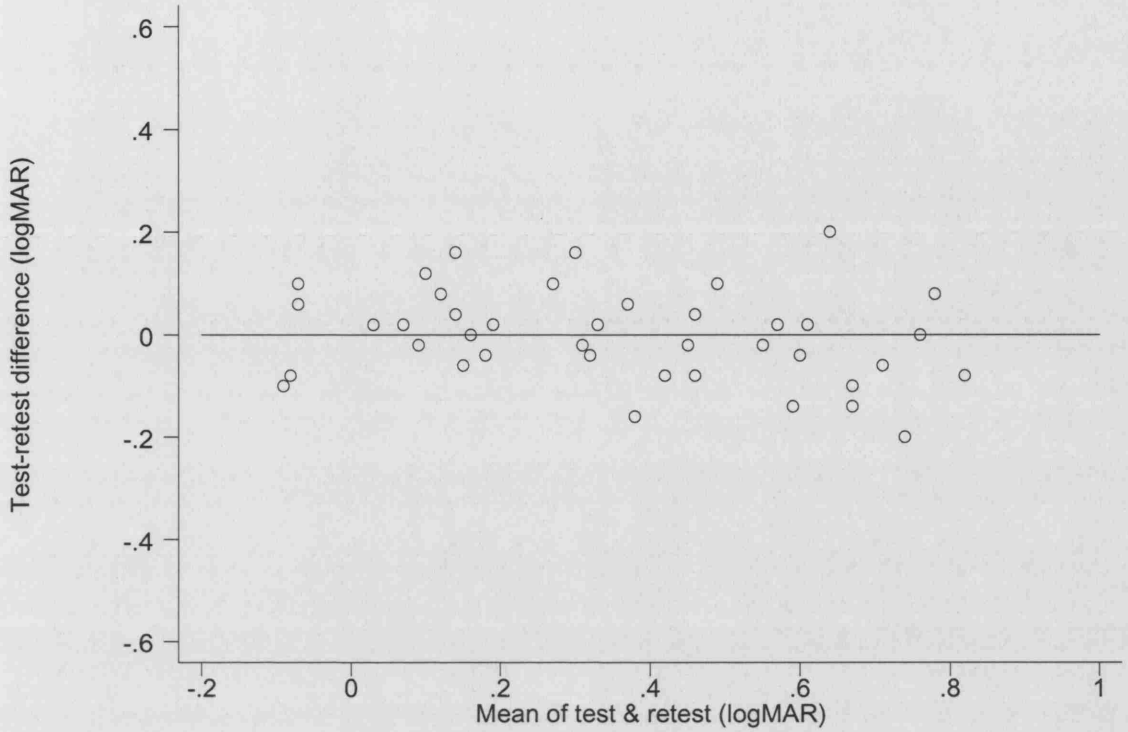


Figure 5.3.8. Bland-Altman plot for precision – RLM-A

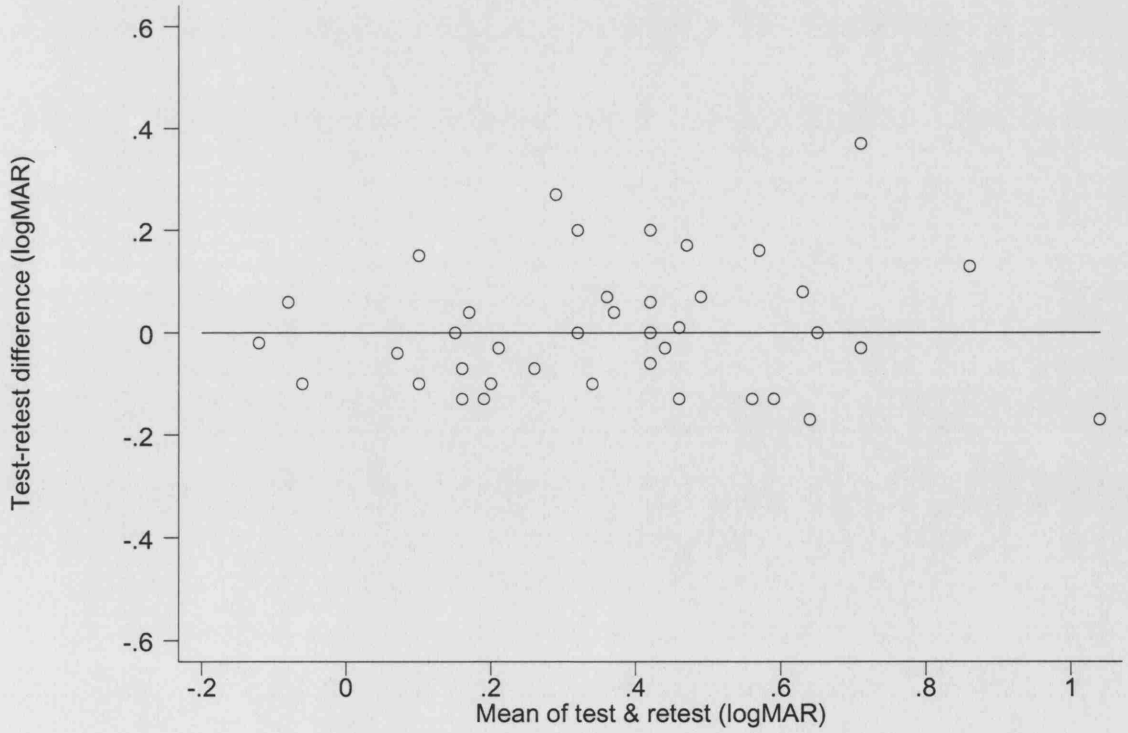


Figure 5.3.9. Bland-Altman plot for precision – RLM-B

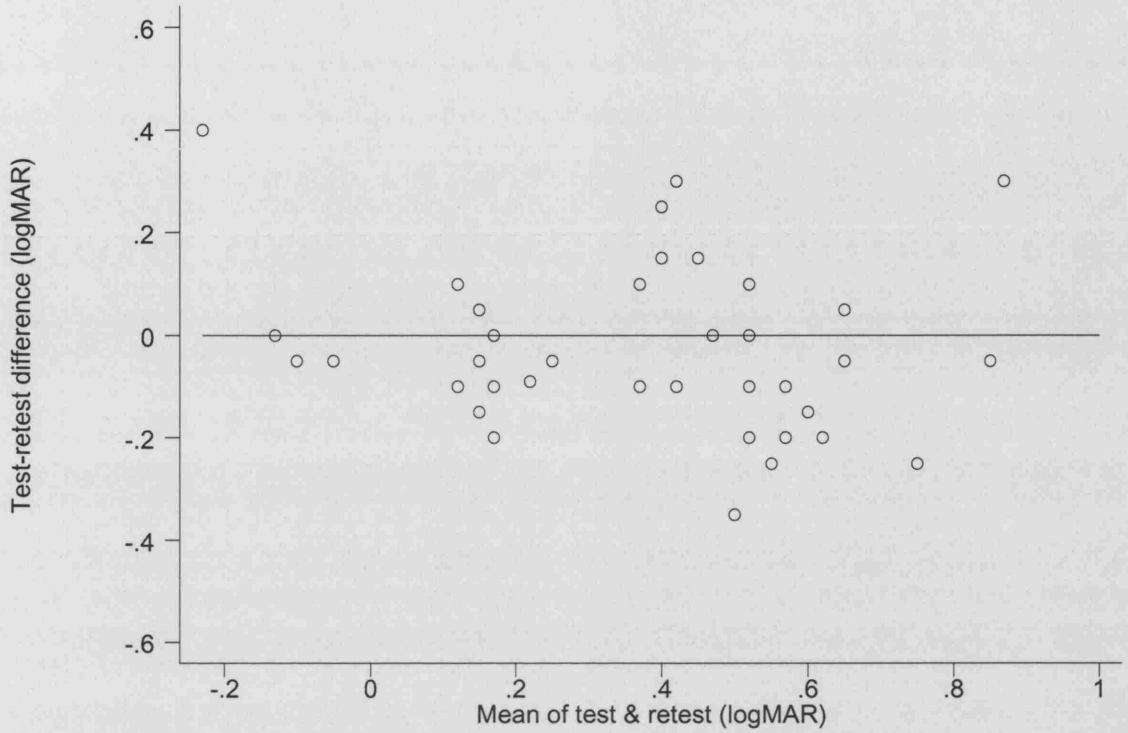


Figure 5.3.10. Bland-Altman plot for precision – RLM-C

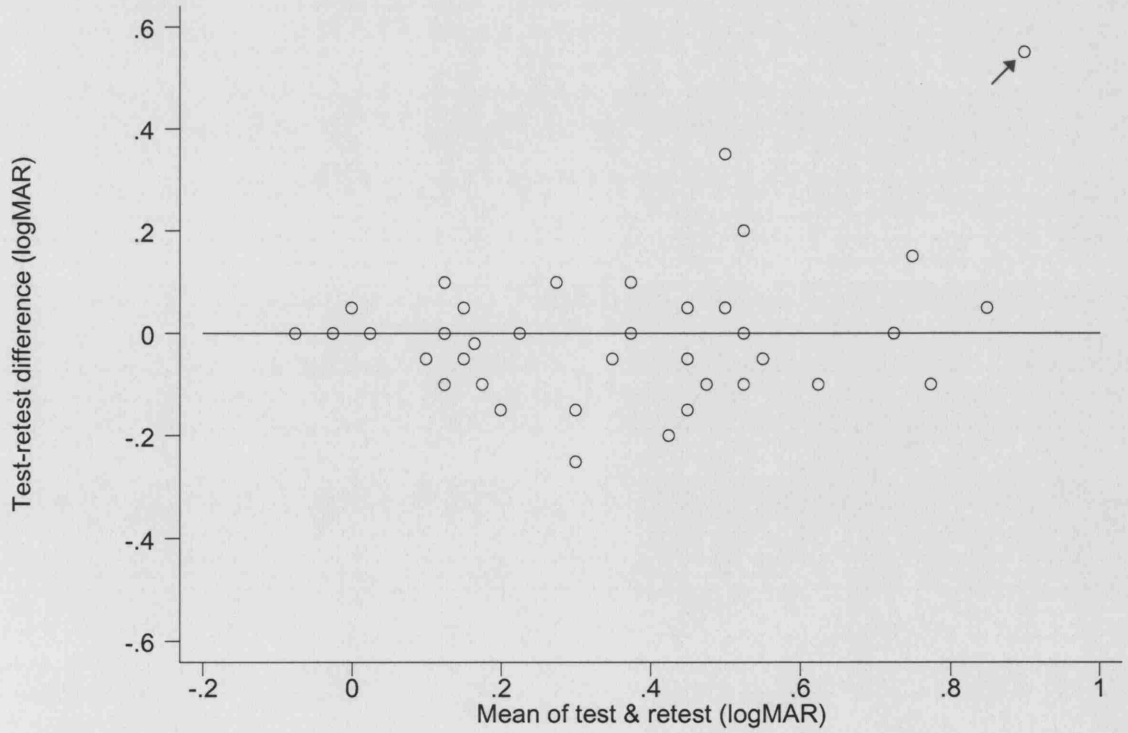


Figure 5.3.11. Bland-Altman plot for Snellen (line-assignment)

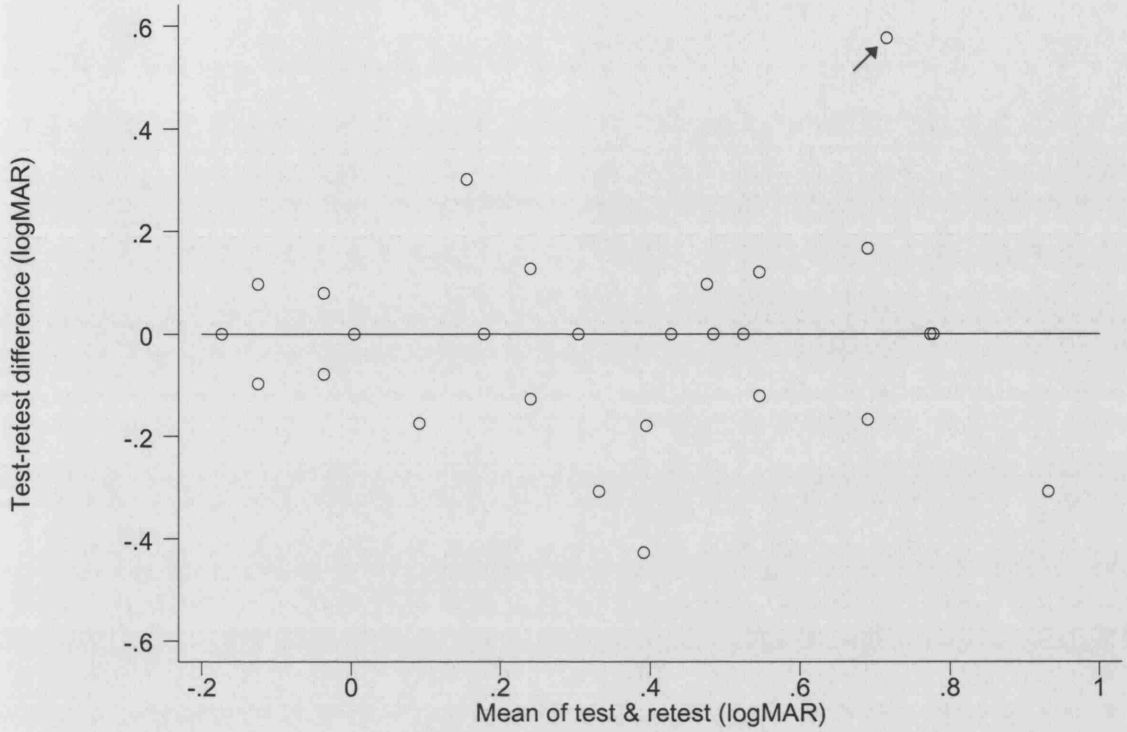


Figure 5.3.12. Bland-Altman plot for Snellen (interpolated)

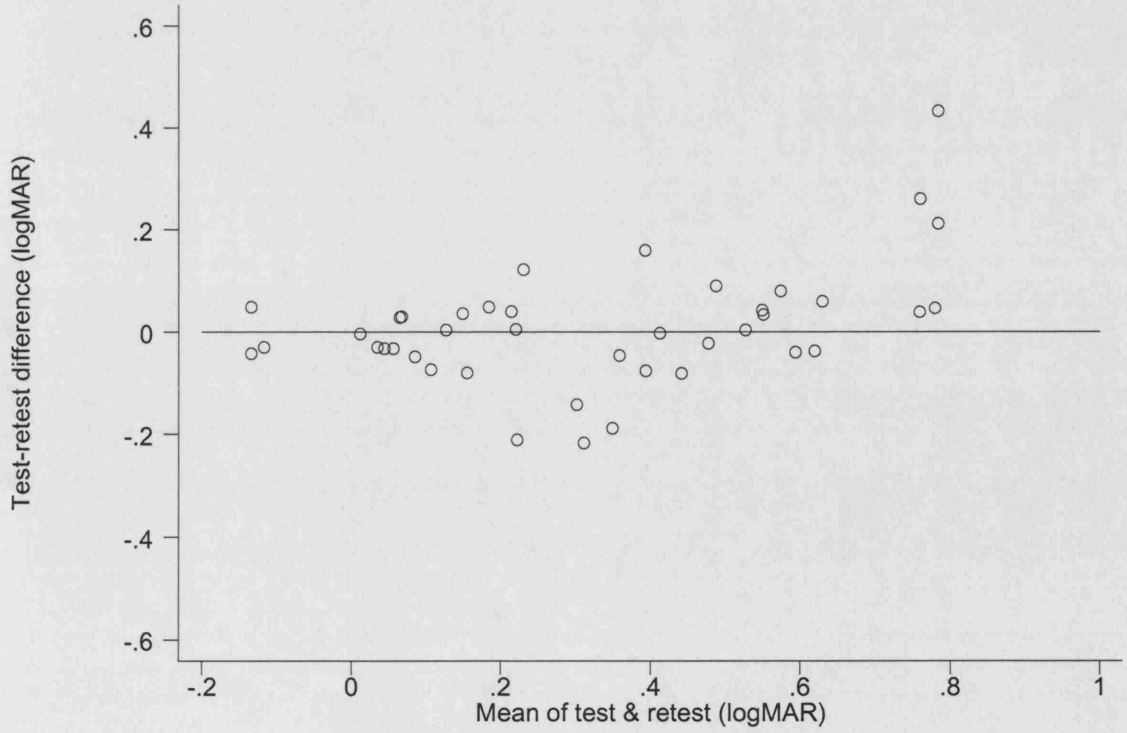


Table 5.3.5. lists estimates of precision for each chart in terms of the 95% TRR (see section 2.3 for definition) for all 41 subjects. The mean difference and its 95% confidence interval give no indication of systematic bias, as would be expected in view of the fact that each pair of measurements was taken on the same design of chart. The ETDRS chart was the most precise (95% TRR ±0.18 logMAR). The RLM-A and Snellen 'interpolated' measurements were less precise, both with a 95% TRR of ±0.24 logMAR; and the RLM-C, RLM-B and Snellen 'line-assignment' acuities were less precise still (95% TRR ±0.27, ±0.31 and ±0.33 logMAR respectively).

Table 5.3.5. Precision of the five chart designs.

Chart	Mean difference (95% CI) (logMAR)	Scale increment (logMAR)	95% TRR (logMAR)	F* (cf ETDRS)
ETDRS	0.00 (-0.03 to +0.03)	0.020	±0.18	-
RLM-A	+0.01 (-0.03 to +0.05)	0.033	±0.24	0.538 (p=0.053)
RLM-B	-0.03 (-0.08 to +0.02)	0.050	±0.31	0.321 (p=0.001)
RLM-C	-0.01 (-0.05 to +0.03)	0.050	±0.28	0.417 (p=0.007)
Snellen (Line-assign')	-0.02 (-0.07 to +0.03)	Variable	±0.33	0.291 (p<0.001)
Snellen (Interpolated)	+0.01 (-0.03 to +0.05)	Variable	±0.24	0.597 (p=0.107)

* - F-test for a difference in variance (see section 4.1.6)

The F-test test (section 4.1.6) was used to compare the precision of each chart with that of the ETDRS chart. The final column of Table 5.3.5 lists the results of this analysis. The variance of the distribution of differences between test and retest was significantly greater for RLM-B, RLM-C, and Snellen (line-assignment scoring) measurements compared with that of the ETDRS chart. The difference between the estimates of precision for the RLM-A and ETDRS charts approached significance at the 5% level (p=0.053), as did the difference between the RLM-A and Snellen (line-assignment) estimates (F=0.553, p=0.065).

To assess the effect of the outlying points identified in Figs 5.3.5 and 5.3.6, the 95% TRR was recalculated for the RLM-C and Snellen charts with these respective points

excluded. The 95% TRR for the RLM-C chart excluding subject 28 was ±0.22 logMAR, and that for the Snellen chart excluding subject 2 was ±0.27 logMAR.

Precision and underlying acuity

The Bland-Altman precision plots for the RLM and Snellen charts (Figs 5.3.6-12) appear to demonstrate a wider vertical spread of points towards the right of the plot. This may be suggestive of a reduction in precision for poorer acuities. Evidence of a relationship between precision and underlying acuity was sought for each chart using linear regression analysis (see section 4.1.6). This was done by regressing the unsigned difference values against the mean of the two measurements (see Table 5.3.6). This analysis suggested a weak relationship between precision and underlying acuity for the Snellen chart (t=2.11, p=0.041), and a stronger relationship for the RLM-C chart (t=2.8, p=0.008). However, this finding was again due in the main to the two outlying points identified in Figs 5.3.5 and 5.3.6 (and highlighted with blue arrows in plots 5.3.10 and 5.3.11). With these points excluded, the relationship was no longer significant at the 5% level for either the RLM-C chart (t=1.55, p=0.129) or for the Snellen chart (t=1.52, p=0.136).

Table 5.3.6. Using linear regression to seek a relationship between precision and underlying acuity

Chart	Slope of regression line
ETDRS	0.03 (t=0.94, p=0.354)
RLM-A	0.09 (t=1.96, p=0.056)
RLM-B	0.07 (t=1.17, p=0.250)
RLM-C	0.17 (t=2.80, p=0.008)
Snellen*	0.14 (t=2.11, p=0.041)

* - scored with line-assignment method

Time taken to do test

Histograms were generated to assess the distribution of chart reading time for each chart design. The histogram for the ETDRS chart is shown in Fig 5.3.13. For all 5 chart designs these histograms were strongly suggestive of a non-normal distribution, each showing a positive skew. Accordingly, a logarithmic (base-10) transformation (see section 4.1.4) was applied to the distribution of time data for each of the 5 tests. The histogram for the transformed data for the ETDRS chart is shown in Fig 5.3.14. The distributions of log-transformed chart reading time were assessed for normality using the Shapiro-Wilk W-test. No evidence of non-normality existed for any of the 5 tests (see Table 5.3.7).

Figure 5.3.13. Histogram of ETDRS chart reading time

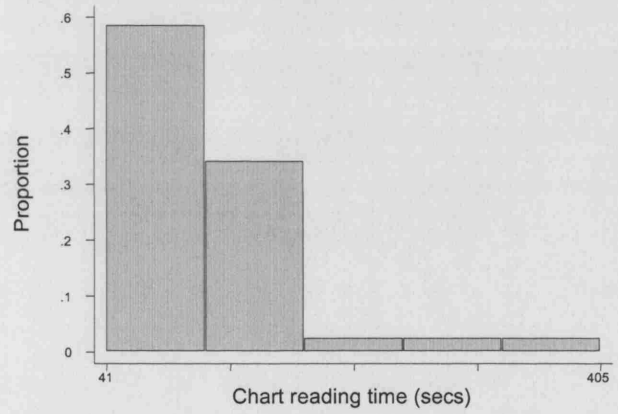


Figure 5.3.14. Histogram of log-transformed ETDRS chart reading time

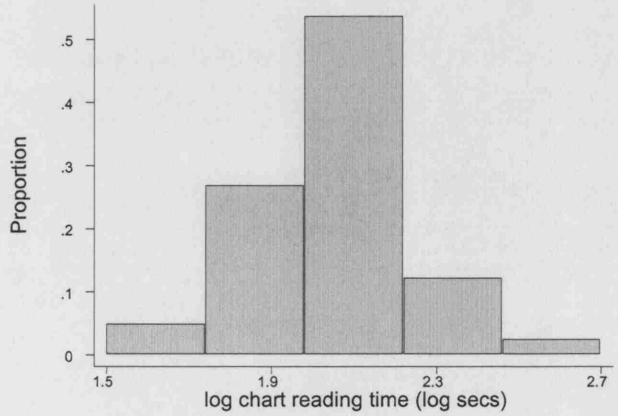


Table 5.3.7. Assessing log-transformed chart-reading time data for normality

Chart	W*
RLM-A	0.98 (p=0.54)
RLM-B	0.98 (p=0.85)
RLM-C	0.97 (p=0.33)
Snellen	0.97 (p=0.45)
ETDRS	0.98 (p=0.71)

* - Shapiro-Wilk test (see section 4.1.6)

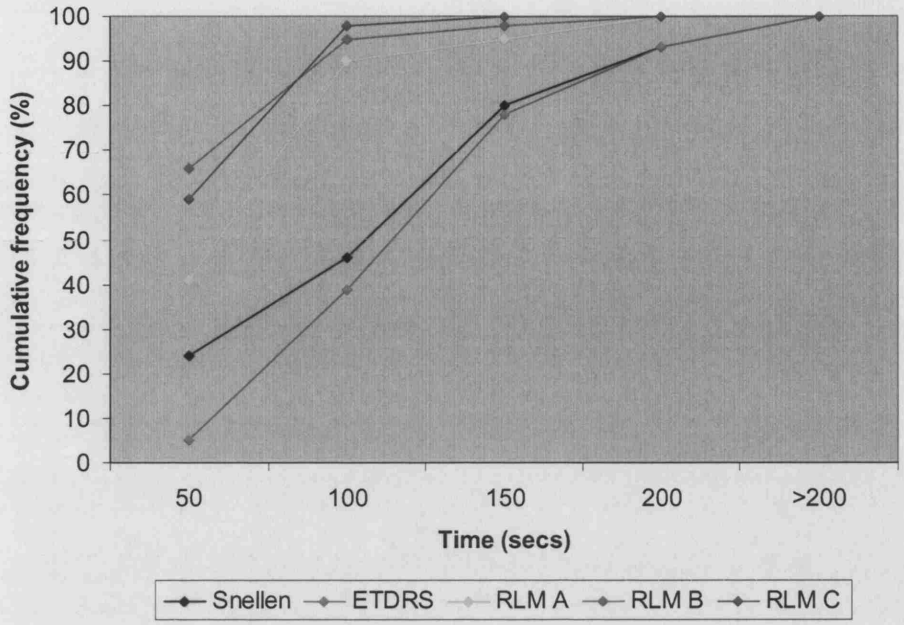
Following transformation back into the original units, the mean chart reading time for the prototype charts is approximately half that of the Snellen and ETDRS charts (Table 5.3.8). Paired t-tests (see section 4.1.6) were performed on the log-transformed data to compare the time taken for a measurement with each prototype chart with that for both the ETDRS chart and the Snellen chart. This showed that measurement time was significantly shorter for each of the prototype charts compared with the ETDRS or Snellen charts. There was no statistical difference between the time required for an ETDRS and a Snellen measurement. The distribution of chart reading time is shown diagrammatically in Fig 5.3.15.

Table 5.3.8. Analysis of chart reading time following $\log_{10}$ transformation

Chart	Total no of letters	Mean$ (secs)	Range (secs)	Paired t-test* (cf ETDRS)	Paired t-test** (cf Snellen)
RLM-A	39	54	13-182	10.79 (p<0.001)	4.52 (p<0.001)
RLM-B	26	40	14-161	13.07 (p<0.001)	6.58 (p<0.001)
RLM-C	27	48	19-104	10.87 (p<0.001)	6.22 (p<0.001)
Snellen	45	93	19-417	1.73 (p=0.091)	-
ETDRS	120	111	41-405	-	1.73 (p=0.091)

$ - mean of log-transformed data
* - paired t-test for difference in means compared with ETDRS chart using log-transformed data
** - paired t-test for difference in means compared with Snellen chart using log-transformed data

Figure 5.3.15. Chart reading times for interpolated measurements



6. DISCUSSION

This study has compared the performance of three prototype logMAR visual acuity charts with that of the Snellen chart, and the gold standard ETDRS chart in subjects with a range of visual acuity. The results suggest that charts which feature an abbreviated version of the ETDRS design produce visual acuity measurements which are accurate compared with those of the gold standard ETDRS chart across a range of acuities. As such, acuities measured using such prototype charts may be compared directly with those measured with the ETDRS chart.

When scored using the line-assignment method with which the chart is typically used, Snellen acuities were, on average, 0.07 logMAR better than ETDRS interpolated measurements. This difference equates to approximately ¾ of a row of ETDRS letters and was statistically significant. The difference was related to the underlying level of acuity such that it was maximum for good acuities and less for acuities towards the poorer and of the range. The observed bias may be due to one or more of a number of potential confounding variables inherent in the Snellen design (e.g. variable number of letters per line, uncontrolled letter legibility, variable letter spacing) and is exacerbated by line-assignment scoring (see Tables 5.3.2 and 5.3.3). The difference in test luminance may have contributed to this bias but is unlikely to be the sole cause as the magnitude of the observed bias would be inconsistent with previous work on the strength of the relationship between visual acuity and test luminance (see section 2.10.9). Also, acuity-related bias caused by higher luminance might be expected to favour poor acuities rather than good acuities as was observed. Regardless of the cause, the observed degree of bias limits the extent to which Snellen acuities can be compared with ETDRS acuities, even after converting the Snellen fraction into logMAR notation.

The precision of an acuity test is important because it determines the ability of the test to detect change (see sections 2.8 and 2.9). Precision is strongly influenced by the scale increment of the test (see section 2.10.6

and Table 5.3.5), which in turn is determined by the number of letters per line and the size interval between lines. An increase in the number of letters per line and/or a reduction in the inter-line size interval will result in increased precision. It also results in a larger total number of letters, thereby increasing the time required for a measurement. A trade-off therefore exists between precision and measurement duration. This study has estimated the precision (in terms of the 95% TRR – see section 4.1.2) of the ETDRS chart to be ±0.18 logMAR. This figure appears in keeping with previously reported estimates made using either the Bailey-Lovie or ETDRS chart which range from ±0.07 to ±0.19 logMAR[74 78-81 84]. The mean time required for an ETDRS measurement was 111 seconds (Table 5.3.6). With a 95% TRR of ±0.24 logMAR, the RLM-A chart was less precise than the ETDRS chart ($F=0.538$, $p=0.053$) but the chart reading time was 54 seconds, approximately half that required for an ETDRS measurement ($p<0.001$). This represents an average time saving of around 1 minute per measurement. Measurements taken with the RLM-B and RLM-C charts required slightly less time (median reading time 40 and 48 seconds respectively), but at the cost of a further reduction in precision (95% TRR ±0.31 and ±0.28 logMAR respectively).

This study estimated the precision of the Snellen chart using the line-assignment scoring method typically employed in clinical practice to be ±0.33 logMAR, significantly worse than for the ETDRS chart ($p<0.001$). The difference between this estimate of precision for the Snellen chart and that for the RLM-A chart was ±0.09 logMAR which approached statistical significance at the 5% level ($p=0.065$). In practice however, the impact of this difference in precision is greater than these figures would suggest, because to find the smallest measured change which can be attributed to a real change in clinical status (see section 2.9) we must look to the next point on the measurement scale. The jump to the next point on a the coarser line-assignment scale is likely to be larger than for an interpolated score[80]. As well as producing the lowest estimate of precision, the Snellen chart required an average of 93 seconds for a measurement. This is

slightly faster than an ETDRS measurement (p=0.091), but significantly slower than an RLM-A measurement (p=0.001).

This study has utilised the methods of Bland and Altman to estimate precision (see section 4.1.2). A prerequisite of the use of these methods is that the distribution of differences from which the estimates are derived is approximately normal (see section 4.1.4). The distributions of test-retest differences for the RLM-C and Snellen charts showed evidence of a departure from normality which for each chart was due in the main to a single outlying point (see Table 5.3.2 and Figs 5.3.5 and 5.3.6). Reanalysing with these points excluded produced a narrower 95% TRR for both charts. The more optimistic precision estimate for the Snellen chart (±0.27 logMAR) was still inferior to that of the ETDRS (±0.18 logMAR) or RLM-A (±0.23 logMAR) charts. The revised estimate for the RLM-C chart (±0.22 logMAR) was slightly narrower than that for the RLM-A chart (±0.23 logMAR), and much narrower than that for the RLM-B chart (±0.31 logMAR). In view of the fact that the scale increment of the RLM-C chart is larger than that of the RLM-A chart, and the same as that of the RLM-B chart (see Table 5.2.1), this revised estimate appears somewhat inconsistent. Interestingly, if we attempt to predict the precision of the RLM charts based upon the empirical estimate of precision for the ETDRS chart (for which the test-retest differences were normally distributed) and a knowledge of the difference in scale increment[A], we obtain figures which are more similar to the original empirical estimates before any points were excluded (see Table 6.1).

[A] Theoretically, we would expect that making the scale of a visual acuity chart X times finer would reduce the standard deviation of the differences between repeated measures by a factor of $1/\sqrt{X}$ (see section 2.10.6).

Table 6.1. Predicting the precision of the RLM charts from an empirical estimate for the ETDRS chart

Chart	Scale increment (logMAR)	Empirical 95% TRR*	Predicted 95% TRR**
ETDRS	0.020	±0.18	-
RLM A	0.033	±0.23	±0.24
RLM B	0.050	±0.31	±0.28
RLM C	0.050	±0.28	±0.28

* - Empirical estimates from present study.
** - Derived from empirical observations for the ETDRS chart (see Table 5.3.5).

For the RLM-C and Snellen charts, there appeared to be a relationship between precision and underlying acuity (see Table 5.3.6). However, the statistical significance of this finding hinged on the same outlying points referred to in the previous paragraph. This prevents any unequivocal conclusions from being drawn from this observation. However, an interesting observation was that both these points related to subjects with relatively poor levels of acuity; subject 28 (ETDRS acuity +0.76 logMAR) and subject 2 (ETDRS acuity +0.59 logMAR). A larger study with subjects stratified into various acuity ranges would be required to provide unequivocal conclusions as to the relationship between precision and underlying acuity.

In summary, ETDRS acuities are relatively precise but time consuming to measure. Whilst less precise, RLM-A acuities can be measured in half the time of an ETDRS measurement, but with greater precision than the Snellen chart (even after allowing for non-normality in the distribution of test-retest data for the Snellen chart). In addition, RLM-A acuity measurements are accurate compared with ETDRS measurements, whereas the Snellen chart exhibits acuity-related bias. In view of these findings, along with the benefits of the design principles of Bailey and Lovie (see section 2.13), the RLM-A chart would appear to be preferable to the Snellen chart as the test of choice in routine clinical practice. The precision of the Snellen chart can be improved through the use of interpolation (see Table 5.3.3), but unlike for the other charts featured in this study, this is

too cumbersome a procedure for routine clinical use[A]. The duration of a Snellen acuity measurement in this study was twice that of a RLM-A measurement. A Snellen acuity measurement could be carried out more quickly (and almost certainly is in routine practice) by avoiding the use of rigorous termination rules and forced-choice methods, but this exposes the resultant acuity score to the influence of various sources of patient- and examiner-related bias (see section 2.10.7).

[A] To obtain a realistic measurement time for an interpolated Snellen measurement, the time required to calculate the score from the raw data must be added to the 'chart reading time' listed in Table 5.3.8. In the author's experience the calculation of an interpolated Snellen score requires approximately 1 minute compared with a few seconds for any of the other charts featured in this study.

The 'RLM-E' chart

The prototype RLM-A chart as described in section 5 was felt to offer certain advantages which would be pertinent to the measurement of visual acuity in large-scale surveys of eye disease and/or visual impairment (see section 2.8). To allow a single chart to be used with any language and degree of literacy, a 'tumbling-E' version of the RLM-A chart was designed. The resultant 'RLM-E chart' was subsequently employed in a series of population-based studies in various parts of the World. Details of the various studies in which the RLM-E chart has been used, along with sample data from these studies in the field are listed in Appendix A.

Link section

The Bailey-Lovie/ETDRS chart design is considered the gold-standard method of measuring visual acuity (see section 2.13). The estimated width of the 95% TRR for the ETDRS chart in this study was ±0.18 logMAR. With respect to the detection of change, this degree of precision is such that only measured changes of 0.20 logMAR (two rows of letters) or more may be attributed to a true change in clinical status. A change in visual acuity of 0.20 logMAR equates to a change in the minimum angle of resolution of approximately 58%. Although this estimate was in keeping with previously published figures, it represents a relatively poor ability to detect change. For example, an identical degree of precision exhibited by a perimeter would require the threshold at a given location to change from 32 dB to 20 dB before the clinician can be sure that the change is not the result of measurement error alone. As well as its significance with respect to the ability to detect change, a low level of precision is also relevant to the design of clinical trials as the lower the degree of precision (i.e. the larger the degree of test-retest variability or TRV) the larger the sample size required in a clinical trial which seeks to demonstrate a given difference between groups with a given degree of statistical power (see section 2.9).

The variability between test and retest which results in such low estimates of precision can be thought of as measurement error, or 'noise' obscuring the 'signal' of the true level of resolving power. A well recognised psychophysical method of improving the signal to noise ratio is that of averaging, in which the influence of random noise on a signal is reduced by repeatedly measuring the signal and averaging the multiple measurements[52]. The ratio of signal to noise will progressively improve as more measurements are averaged on account of the random nature of the noise. It would seem reasonable to suggest that repeating and averaging may provide a means of improving the precision of visual acuity measurements. A personal computer would have the potential to carry out repeating and averaging of acuity measurements as well as displaying the visual acuity stimuli themselves.

In proposing a replacement for the current gold standard which utilises the same acuity notation, it is desirable for the test to produce acuity measurements which are accurate when compared with those of the existing gold standard, such that acuities measured on both are comparable. The design of the test and the scoring method incorporated in the software, were therefore required to match that of the ETDRS chart as closely as possible. Another reason for developing a test specifically for the purpose was to enable the manipulation of certain test parameters such that their effect upon test performance may be elucidated. These requirements were not met by any existing test, therefore a novel test was developed.

7. DEVELOPING A COMPUTERISED SYSTEM TO MEASURE AND AVERAGE REPEATED ACUITY THRESHOLDS

7.1. AIM

- To develop a test capable of conducting repeated measurements of visual acuity and calculating their average.

7.2. OVERVIEW OF COMPUTERISED TEST DESIGN (THE 'PC-TEST')

7.2.1. *Equipment*

Hardware

A 20-inch cathode ray tube display (Ergovision 2040, Taxan Europe Ltd) was employed to display the visual acuity stimuli. This screen had a native resolution of 1024x768 pixels combined with a non-interlaced refresh rate of 85 Hz. A standard IBM compatible 150 MHz PC was used to drive the monitor as well as to receive examiner responses and perform any required data analysis. This equipment will be referred to throughout as the 'PC-test'.

Software development

Software was commissioned via a collaboration with the Laboratory of Physiological Optics, Institute of Ophthalmology, University College London. 'Acuity Analyser' was a dedicated programme developed to meet the research group's design specifications. Software was written by a team familiar with the development of equipment for the purpose of presenting psychophysical stimuli for vision research[A].

7.2.2. *PC-test software – 'Acuity Analyser'*

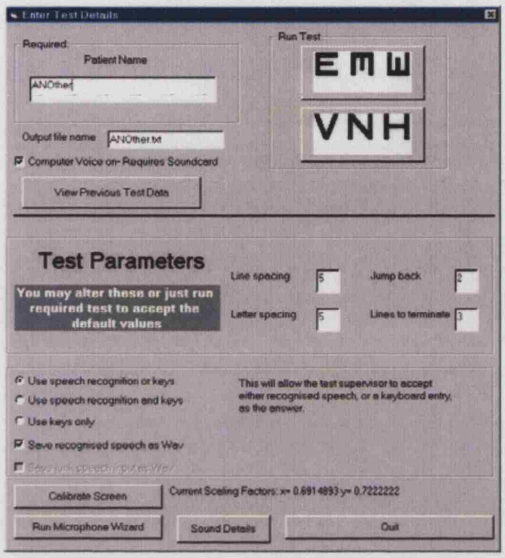
Input screens

On initiation of the Acuity Analyser software the operator is prompted to input their name. The input of a password at this stage gives the operator privileges which will allow them to enter the calibration section of the software. This is necessary the first time the system is used with a given display monitor. Fig 7.2.1 shows the Acuity Analyser input screen with and without calibration privileges. Initiation of the calibration process presents an input screen featuring a box as shown in Fig 7.2.2. To ensure that the test stimuli are scaled appropriately for a particular display, the horizontal and vertical dimensions of the box as displayed on the screen are entered in the appropriate boxes. These values are saved and retained until the calibration process is repeated.

[A] The copyright for this software is retained by the Laboratory of Physiological Optics, University College London. The details of this software were not disclosed for publication in this thesis.

Figure 7.2.1. 'Acuity Analyser' Input Screen a) with calibration privileges, b) without calibration privileges

a)



b)

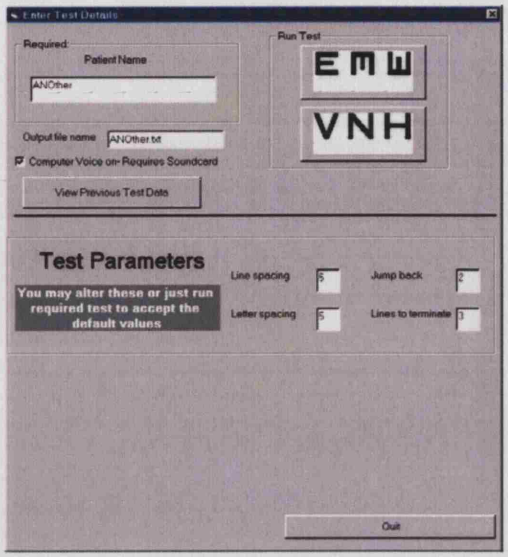
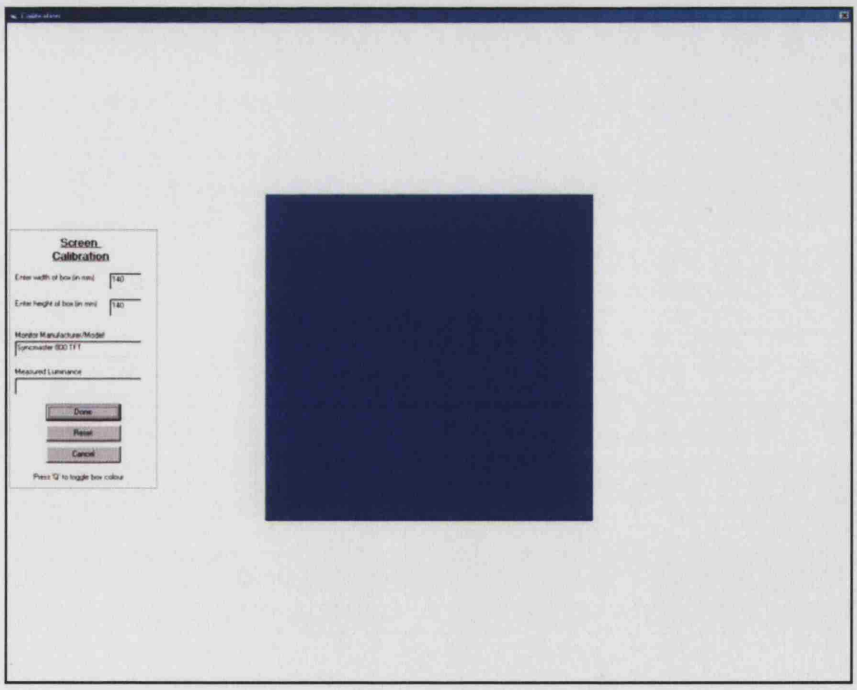


Figure 7.2.2. Calibration screen



Required information for the test details input screen includes the following (described in more detail below):

- Type of optotype (Sloan letters or Tumbling-E)
- Patient name (which is used to create the file name for the output of test results)
- Line spacing (in stroke widths where a letter is based on a 5x5 grid)
- Letter spacing (in stroke widths)
- Jump (see section 7.2.3)
- Number of thresholds (see section 7.2.3)

Test screen

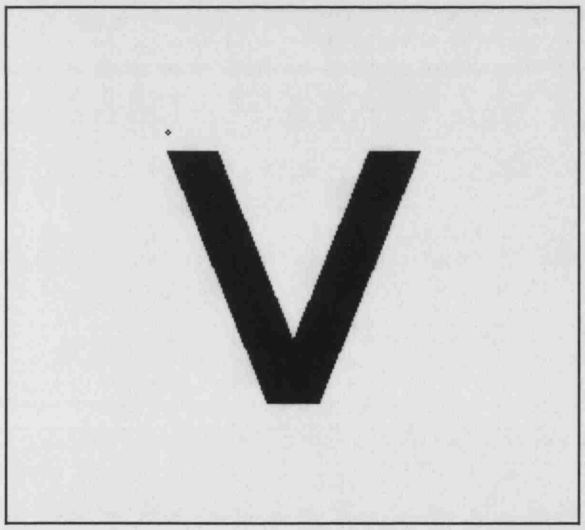
The arrangement of test stimuli was designed to mimic that of the reference standard ETDRS logMAR chart as closely as possible. The letter stimuli used were those of the Sloan letter set[98] (see section 2.13) and were drawn on a 5x5 grid exactly to Sloan's specifications (Fig 2.15.1). The display featured 3 horizontal rows of 3 letters, the centre row of the three containing the three test stimuli.

Figure 7.2.3. Sample PC-test display

The upper and lower rows of letters were included to simulate the contour interaction provided by surrounding letters on the ETDRS chart (see Fig 7.2.3). Letter quality was

maximised by employing anti-aliasing[A]. The number of stimuli per row was limited to 3 due to the constraints of the display. Otherwise, the default display format followed that of the ETDRS convention, with inter-letter spacing of 5 stroke widths (equivalent to the width of one letter on the 5x5 grid), and an inter-line spacing of 5 stroke widths, where the stroke width is that of the letters in the upper line. The size of the letters in the upper line was fixed at 0.1 logMAR larger than that of the test stimuli, and that of the lower line was fixed at 0.1 logMAR smaller. The letters were generated at random, the only restriction being no repeats on a given line. The test display also included a small marker to indicate to the examiner which is the next letter to be read by the subject (Fig 7.2.4). The marker was designed such that it was not visible to the subject during testing. Background luminance was measured at 90 cd/m^2 with a contrast of 92%.

Figure 7.2.4. Stimulus close-up



Marker of letter to be read can be seen near top left corner of stimulus

[A] This is a technique whereby contrast at the edge of the optotype is modified to prevent departures from the intended stimulus specification occurring due to the spatial limitations of a pixilated display.

7.2.3. *Test paradigm*

Each PC-test measurement consists of a series of individual acuity thresholds. Each measurement commences at a stimulus size of +0.80 logMAR. The examiner enters the subject's responses to the three stimuli in the central row at this size level. Providing at least one of the three responses is correct, the stimulus size decreases progressively by 0.1 logMAR. This represents a 'scrolling' of the chart down to the next line such that the dimensions of stimuli, surrounding optotypes and the spaces between them all decrease by 0.1 log units. Once the subject has made a full line of incorrect responses, that threshold is terminated and the next in the series initiated (unless that threshold is the last in the series). The initial stimulus size for the next in the series of thresholds is determined by adding a predetermined value to the final stimulus size of the preceding threshold (up to a maximum of +0.80 logMAR). This value is specified by the operator at the beginning of the test and is referred to as the 'jump'. For example a jump of 3 indicates a 3 line (or +0.30 logMAR) increase in stimulus size between the end of one threshold and the beginning of the next. The process is then repeated for the next in the series of repeated thresholds. The total number of thresholds is also specified prior to the test. At the termination of each measurement the programme creates an output file containing details of the test (see Fig 7.2.5). These include the size of each group of three stimuli, the number of errors per group of three stimuli, the individual threshold acuity scores, and their mean. The date, time and duration of the test is also detailed in the output file.

7.2.4. *Scoring method*

To score each of the multiple acuity thresholds with a single measurement, the PC-test uses a scoring method which is equivalent to the interpolated scoring method used with the ETDRS chart. When using the ETDRS chart, the final acuity score is typically derived from the total number of letters correctly named (see section 4.2.9). An adjustment to the method is required for the PC-test as the size of the initial stimulus will vary between thresholds. Therefore the score is derived from the last row upon which a correct response was made, and 0.03333 logMAR is added to this value for each misnamed letter up to that point. At the end of the test, the individual interpolated acuity scores for each threshold are averaged to produce the final acuity score.

Psychometric scoring methods have been used in computerised visual acuity systems[190]. Such an approach would seem more suitable for computerised tests than for standard acuity chart measurements as the required analysis can be fully automated. Vanden Bosch and Wall[78] have suggested that for charts which feature more than 5 letters per line, psychometric scoring may offer improved precision over that achieved with interpolated scoring[A]. This has since been shown not to be the case at least for charts with eight letters per line[141]. In the absence of any evidence of improved precision when using psychometric scoring methods, interpolated scoring was used with the PC-test to promote good agreement with the ETDRS chart.

[A] For the PC-test as described herein, the total equivalent number of letters per line can be found by multiplying the number of letters per line during each individual threshold (3) by the total number of repeats (variable).

Figure 7.2.5. Sample Output File (section removed)

```
-------------------------------------------------
|CustTest - Chart presentation program with staircase  |
|Copyright Inst. Opthalmology/F.W. Fitzke and A.Wade '98|
-------------------------------------------------
Thursday, October 15, 15:35:37

Trial 0
LineSize=1          SCORE 0
Trial 1
LineSize=2          SCORE 0
Trial 2
LineSize=3          SCORE 0
Trial 3
LineSize=4          SCORE 0
Trial 4
LineSize=5          SCORE 0
Trial 5
LineSize=6          SCORE 0
Trial 6
LineSize=7          SCORE 0
Trial 7
LineSize=8          SCORE 0
Trial 8
LineSize=9          SCORE 2
Trial 9
LineSize=10         SCORE 3
Trial 10
LineSize=5          SCORE 0
Trial 11
LineSize=6          SCORE 0
Trial 12
LineSize=7          SCORE 0
Trial 13
LineSize=8          SCORE 1
Trial 14
LineSize=9          SCORE 0
Trial 15
LineSize=10         SCORE 2
Trial 16
LineSize=11         SCORE 3
Trial 17
LineSize=6          SCORE 0
Trial 18
.
.
.
LineSize=10         SCORE 2
Trial 34
LineSize=11         SCORE 3
Test Completed
Termination line numbers : 10 , 11 , 9 , 10 , 11 ,
TIME
Thursday, October, 15, 15:41:00
Elapsed time: 0 00:05:23

---------------------------------------
TEST PARAMETERS
# Of lines wrong for termination :5
Lines to jump back after wrong line :5
Letter Spacing (in strokes) :0
Line Spacing (in strokes) :0
SCORES
0.067
-0.033
0.100
0.067
-0.033
SumX = 0.167
SumXSQ = 0.021
JustRight=5
Mean acuity=0.033
SD score acuity=0.062
SEM score = 0.028
Comments:

---------- END OF FILE ---------
```

7.3. ESTABLISHING THE OPTIMUM NUMBER OF THRESHOLDS AND THE OPTIMUM 'JUMP'

7.3.1. *Aim*

- To estimate the optimum values for two key PC-test parameters: the number of thresholds and the 'jump' (see 7.2.3 for definitions).

7.3.2. *Methods*

The principle behind the pilot study was to take a series of measurements with a large number of thresholds, and a large jump value. The results can subsequently be analysed in such a way as to determine the likely effect of using a small number of thresholds, or a smaller 'jump'. The standard error of the individual acuity thresholds within a PC-test was used to predict the likely relationship between number of thresholds and precision. 12 subjects were recruited from the patients and relatives attending an ophthalmic outpatient clinic. Diagnoses were as shown in Table 7.3.1. One acuity measurement was taken on each subject using the PC-test as described in section 7.2. The number of thresholds was set at 10 for each PC-test measurement. In theory, this number of repeated thresholds results in a scale increment equivalent to that of an ETDRS chart with 30 letters per line. It was thought unlikely that an increased number of thresholds over and above 10 would offer an improvement in precision. The 'jump' (see section 7.2.3 for definition) was set at 7 lines of letters. This value was thought to be as large as was practical considering the limits of the range of the test at a 4 metre testing distance.

Table 7.3.1. Subject diagnoses

Diagnosis	Frequency
Normal	5
Pseudophakia	3
Uncomplicated LASIK (Laser in-situ keratomileusis)	1
Corneal haze post photo-refractive keratoplasty	1
Cataract	1
Rejecting penetrating keratoplasty	1

7.3.3. *Results*

Optimum number of thresholds

Median acuity with habitual correction was +0.01 logMAR (range -0.22 to +0.40 logMAR). For subject 1, the standard error (SE) of the mean of the series of individual thresholds which make up the PC-test measurement was calculated. This was first done using only the first two of the ten thresholds. SE was then recalculated for the first three thresholds, followed by the first four thresholds, and this process was repeated up to the full ten thresholds. This was done for each subject giving nine SE values for each subject. The SE values corresponding to the mean of the first two thresholds were then averaged across all subjects. This was then done for the first three thresholds, the first four, etc up to ten. The average of the SE values was then plotted against the number of thresholds (Fig 7.3.1). It can be seen that SE reduces with increasing number of thresholds. The curve flattens somewhat after 4 thresholds, but the SE continues to reduce up to and including 10 thresholds.

Optimum 'jump' (see section 7.2.3 for definition)

Fig 7.3.2 shows the effect upon the acuity score of taking account of errors on a progressively greater number of lines prior to the final row of each acuity threshold. The reference value is the acuity score calculated taking account of any mistakes on the last seven lines of each acuity threshold. The graph shows that mistakes on at least the final five rows of an acuity threshold should be taken into account when the acuity score is calculated. If mistakes are only taken into account on the final four rows or less, then the acuity score tends to improve. This suggests that for this group of subjects, errors are unlikely any earlier than 5 rows before the end of the threshold.

Figure 7.3.1. The relationship between number of thresholds and the standard error of their mean

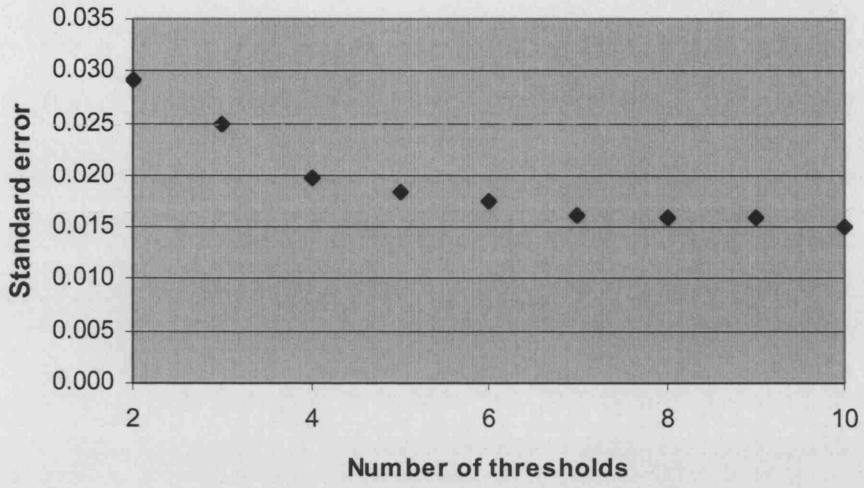
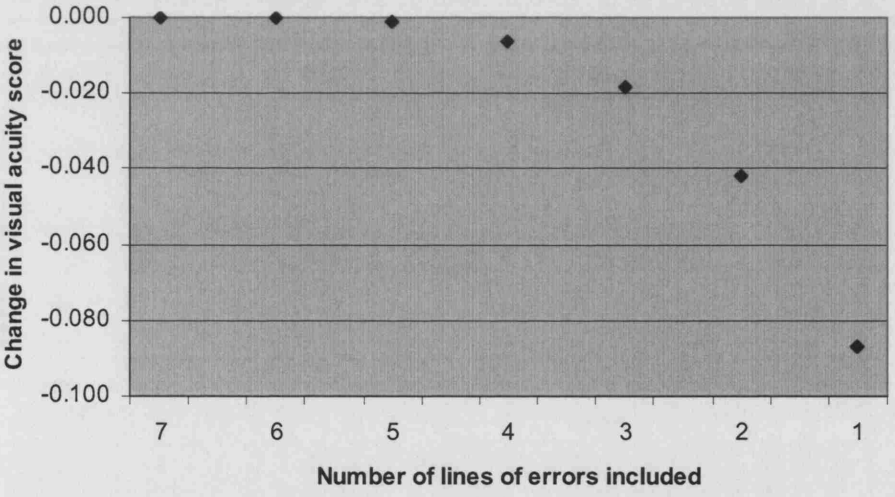


Figure 7.3.2. The effect of accounting for errors on previous lines upon the visual acuity score



7.3.4. *Discussion*

An inadequate number of thresholds within a PC-test measurement would be expected to result in sub-optimal precision, whereas an excessive number of repeated thresholds may increase the total test time without offering any meaningful improvement in test precision. This pilot used the standard error of the individual acuity thresholds within a PC-test measurement to predict the relationship between the number of thresholds within a PC-test and its precision. Increasing the number of individual thresholds within a PC-test measurement reduces the standard error of those thresholds. The results show a pattern of diminishing returns as would be expected from the relationship between the standard deviation and standard error:

$$\text{Standard error} = \sigma / \sqrt{n}$$

Where σ is the standard deviation, and n denotes the sample size. Assuming that SE is a good predictor of precision, the results suggest that a PC-test measurement should contain at least 4 individual acuity thresholds. A greater number of thresholds may improve precision further but the gain may be small and should be weighed against the increased time required for a measurement. Theoretically, the performance of a PC-test which features 3 letters per line and which averages two thresholds of acuity, should equate to that of a single acuity test which features the same inter-line size interval and six letters per line (one more than the ETDRS chart). In view of this, an interesting observation was that the standard error of the mean of a series of thresholds appears to reduce considerably when the number of thresholds per test is increased beyond two. This might be expected to translate into improved precision over that of the ETDRS chart. It should be noted however, that the series of standard errors for each subject in this pilot study are not independent.

Using too small a 'jump' setting would be expected to bias the acuity score in a negative direction (i.e. towards a better acuity score), as the subject has not been given the opportunity to attempt some letters for which they have a high chance of making an error. Conversely, an excessively high 'jump' setting will result in the subject reading a considerable number of letters for which there is a negligible chance of an error, which will result in an unnecessarily long test time. Fig 7.3.2 suggests that a

jump of less than 5 will result in an overestimate of acuity. Another reason for avoiding too low a jump setting relates to the fact that that many subjects find a long series of near threshold stimuli less tiring when punctuated by some easier stimuli[190]. The author's experience with the PC-test would tend to support this conclusion.

A potential criticism of the estimate of the required 'jump' value in this pilot study relates to the potential for ceiling effects. For those with poorer acuities, a large jump value may result in the test's ceiling of +0.80 logMAR being reached. In an attempt to rule out such an effect the results were reanalysed excluding the two subjects with the poorest levels of acuity (+0.40 logMAR and +0.36 logMAR). The resultant median acuity was -0.05 logMAR (range -0.22 to +0.28 logMAR). This reanalysis had no effect on the appearance of Fig 7.3.2 suggesting that ceiling effects did not have a major influence on the results.

Link section

The results of a pilot study suggested that visual acuity may be measured using a computerised system of the type described in section 7.2 and referred to herein as a 'PC-test'. The results suggest that a PC-test measurement which averages at least 4 repeated acuity thresholds (using three letters per line and a 0.1 logMAR inter-line size increment) may offer improved precision over an ETDRS logMAR acuity measurement. It appears that increasing the number of repeated thresholds within a PC-test measurement to 10 may offer a further improvement in precision, although any such improvement is likely to be small. The results also suggest that the parameter referred to as the 'jump' (see section 7.2.3 for definition) should be set to at least 5 lines to avoid overestimating the acuity.

The PC-test as described in section 7.2 and assessed in the subsequent pilot study was deemed a suitable tool for the purpose of investigating whether the average of a series of repeated acuity thresholds is more precise than a single measurement taken with the gold standard ETDRS chart.

8. IS THE AVERAGE OF A SERIES OF REPEATED ACUITY THRESHOLDS MORE PRECISE THAN A SINGLE ETDRS logMAR ACUITY MEASUREMENT?

8.1. AIMS

- To assess the accuracy of the PC-test compared with the ETDRS chart.
- To compare the precision of the PC-test with that of the ETDRS test over two visits.
- To determine whether increasing the number of repeated thresholds within a PC-test measurement from 5 to 10 offers any improvement in precision.
- To determine the length of time required for a PC-test measurement.

8.2. METHODS

8.2.1. *Subjects*

Subjects who fulfilled the following inclusion criteria were recruited from an ophthalmic outpatient clinic:

- able to understand and comply with the testing protocol, and
- stable visual acuity determined via assessment of current clinical diagnosis.

One eye of each subject was assessed (section 4.2.6). Where both eyes met the criteria, the eye with the poorer acuity was used as the study eye.

8.2.2. *Equipment*

ETDRS logMAR chart

The ETDRS chart is described in full in section 2.13. The charts were back-lit in the standard Lighthouse box achieving a luminance of 300 cd/m^2. and contrast of 98%. Versions 1 and 2 of the chart were used.

PC-tests (PC10-test and PC5-test)

The PC-test is described in section 7.2. Two versions of the PC-test were used differing only in terms of the number of repeated thresholds within a single PC-test measurement. The details of the tests are summarised in Table 8.2.1. A PC-test with 5 thresholds (PC5-test) was included based upon the results of the pilot study (see section 7). A PC-test with 10 thresholds (PC10-test) was also included such that the effect upon precision if increasing the number of thresholds could be estimated using independent data sets (which was not the case in the pilot study). A 30 minute warm-up period was employed prior to collection of any data to allow screen luminance to stabilise. The choice of 30 minutes was based upon the results of earlier psychophysical investigations using an identical model of cathode ray tube display[A], and is consistent with the findings of at least one other group using similar equipment[210]. Following this period screen luminance was measured at 90 cd/m^2 with a contrast of 92%.

[A] Unpublished data, Laboratory of Physiological Optics, University College London, London, UK.

Table 8.2.1. Summary of the tests under comparison

Test	Letters per line	Inter-line increment (logMAR)	Thresholds*	Jump*	Scale increment** (logMAR)
ETDRS	5	0.10	1	N/A	0.020
PC10-test	3	0.10	10	5	0.003
PC5-test	3	0.10	5	5	0.007

* - see section 7.2.3 for definitions

** - Scale increment = $\frac{\text{(Inter-line increment)}}{\text{(No of letters per line) x (No of thresholds)}}$

8.2.3. *Scoring*

ETDRS chart

The ETDRS chart was scored using the interpolated scoring method (see section 4.2.6). An endpoint of a full row of errors was used (see section 4.2.6).

PC-tests

The algorithm for these tests is described in detail in section 7.2. The algorithm employs an identical scoring method and termination rule to that used with the ETDRS chart.

8.2.4. *Investigations*

Four acuity measurements were taken on one eye of each subject using ETDRS charts 1 and 2, as well as the PC10-test and PC5-test. The testing order was randomised. Following an interval of not less than two weeks, the subjects attended for a second visit during which they underwent repeat testing using the same four tests, again in random order. All subjects wore their habitual spectacle correction and viewed the acuity tests from a distance of 4 metres.

8.3. RESULTS

Of 35 subjects examined at the first visit, 19 subjects returned for the second visit. Table 8.3.1 shows the diagnoses for all subjects who attended the first visit, along with the diagnoses for those who attended both visits. The median acuity (as measured with the ETDRS chart) for these 19 subjects was +0.12 logMAR (range -0.20 to +0.64 logMAR). On reviewing the raw data it was noted that there was a missing observation from the series of retest data for subject 17 (i.e. the second visit) for the PC5-test. This was due to a computer error in which an output file was not created following one of the PC5-test measurements. For this reason, the analysis includes 18 test-retest pairs for the PC5-test.

Table 8.3.1. Subject diagnoses

Tests	**Frequency**	
	Visit 1	**Visit 1&2**
Glaucoma only	9	6
Glaucoma & cataract	5	1
Cataract only	4	3
Ocular hypertension	2	1
Pseudophakia	2	0
Retinal detachment (repaired)	1	1
Pseudo-exfoliation	1	0
Glaucoma & posterior capsular thickening	1	1
Sub-acute angle closure (resolved)	1	1
Posterior capsular thickening	1	1
Radial keratotomy	1	1
Macular degeneration	1	1
Macular degeneration & cataract	1	0
Normals	5	2
Total	**35**	**19**

Accuracy

The accuracy of the PC-tests as compared with the reference standard ETDRS chart was assessed using the methods of Bland and Altman (section 4.1). The data from the first visit only were used for this purpose as this enabled data from all 35 subjects who underwent measurements at the first visit to be utilised. Table 8.3.2 shows the mean difference between measurements taken with the PC-tests and those measured using the ETDRS chart. For both comparisons (PC10-test versus ETDRS, and PC5-test versus ETDRS) the mean difference is greater than zero, suggesting a tendency for the ETDRS chart to produce slightly more negative (better) acuities than the PC-tests. Although consistent, the magnitude of this bias is relatively small (less than 1½ ETDRS letters) and in both cases, the lower limit of the 95% confidence interval borders on, but does not exclude zero. The similarity between the mean differences for the two PC-tests suggests that increasing the number of thresholds from 5 to 10 does not affect the absolute acuity value.

Table 8.3.2. Accuracy as compared with reference standard ETDRS chart

Tests	Mean difference (logMAR)	95% CI (logMAR)
PC10-test – ETDRS-1	+0.024	-0.001 to +0.049
PC5-test – ETDRS-1	+0.027	0 to +0.054

Acuity-related bias

The data from all 35 subjects from the first visit were used to assess for any evidence of acuity-related bias. Bland-Altman plots (see section 4.1) were created for both pairs of tests (Figs 8.3.1-2). An inspection of these plots shows more points falling above the horizontal midline than below. This is consistent with the small degree of bias detailed in Table 8.3.2. There was no obvious relationship between the degree of bias and the underlying level of acuity. The absence of a statistically significant relationship between the degree of bias and the underlying acuity was confirmed by linear regression (PC10-test: t=-0.59, p=0.56; PC5-test: t=0.91, p=0.37).

Figure 8.3.1. Bland-Altman plot for PC10-test versus ETDRS

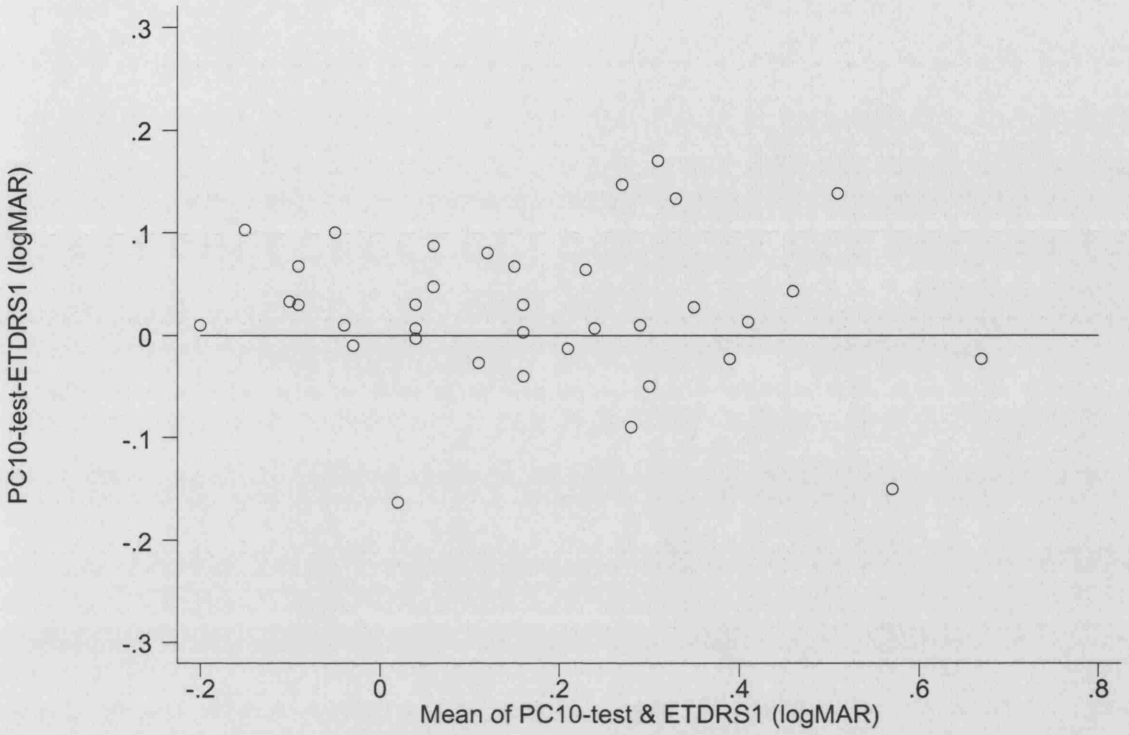
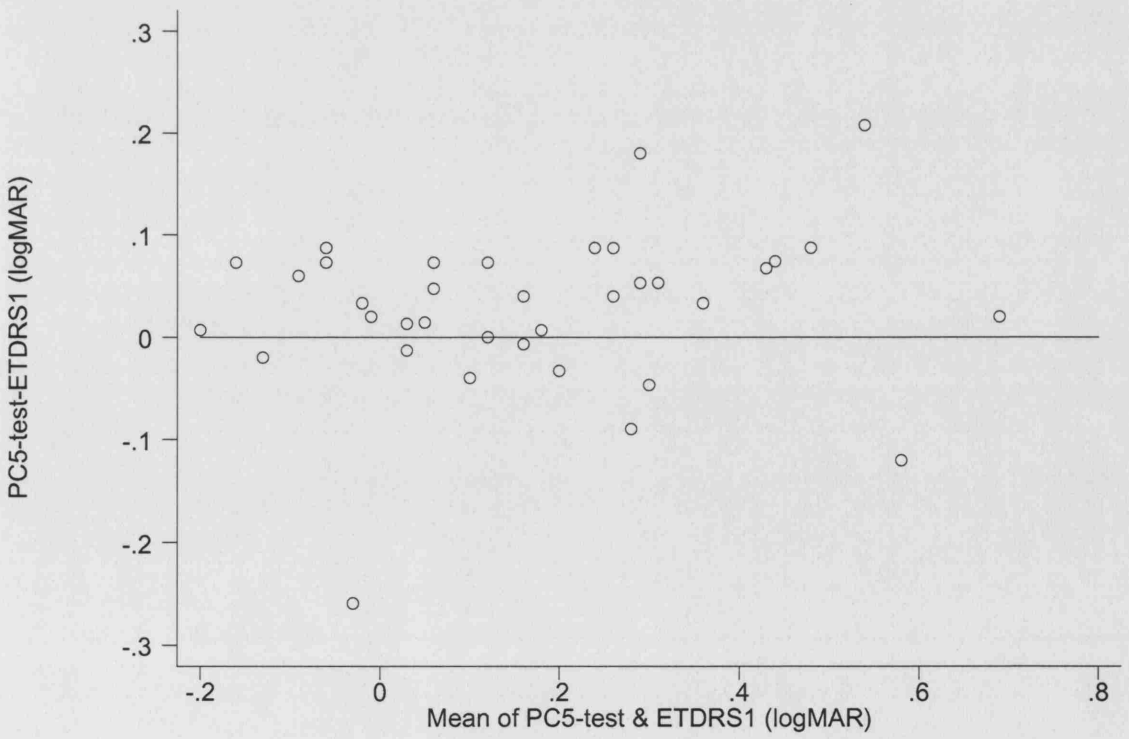


Figure 8.3.2. Bland-Altman plot for PC5-test vs ETDRS



Precision (over two visits)

Estimates of precision were derived from paired measurements, one of which was taken on visit 1 and the other on visit 2. This analysis therefore featured those 18 subjects for whom there was a full set of data for both visits. The Shapiro-Wilk W-test (see section 4.1.5) was used to assess whether the distribution of differences between test and retest conformed to a normal distribution. There was no evidence at the 5% level for a departure from normality for any of the tests (Table 8.3.3).

Table 8.3.3. Assessing test-retest data for departure from normality

Test	W*
PC10-test	0.961 (p=0.592)
PC5-test	0.903 (p=0.064)
ETDRS 1	0.930 (p=0.171)
ETDRS 2	0.910 (p=0.073)

* - *Shapiro-Wilk W-test*

Accordingly, the methods of Bland & Altman were used to estimate precision in terms of the width of the 95% TRR (see section 4.1.2). The widths of the 95% TRR for the PC10- and PC5-tests were ±0.11 and ±0.10 logMAR respectively. As measurements using both versions of the ETDRS chart (versions 1 and 2) were carried out on both visits, estimates of precision were derived from the pooled differences for both versions of the chart. The 95% TRR for the pooled data from ETDRS charts 1 and 2 was ±0.18 logMAR (Table 8.3.4). This represents an increase in the width of the 95% TRR of over two thirds of one ETDRS line (3.5 to 4 ETDRS letters) compared with the PC-tests. The null hypothesis that the variance of the pooled ETDRS differences equalled that of the PC-tests was tested for each PC-test using the F-test (see section 4.1.6). The results of this analysis are also shown in Table 8.3.4 and suggest that the ETDRS chart is significantly less precise than either the PC10-test or the PC5-test.

Table 8.3.4. Test precision

Test	No of thresholds	95% TRR (logMAR)	F** (cf ETDRS)
ETDRS*	1	±0.18	-
PC10-test	10	±0.11	0.532 (p=0.014)
PC5-test	5	±0.10	0.294 (p=0.005)

** - For pooled data for ETDRS versions 1 & 2*
*** - F-test (see section 4.1.6)*

To look for any relationship between the level of precision and the underlying acuity, Bland-Altman plots (see section 4.1) were created from the distributions of test-retest differences for each test (Figs 8.3.1-4). Inspection of these plots did not suggest any relationship between the degree of precision and the level of underlying acuity. Linear regression analysis of the unsigned differences between paired measurements against their mean gave no evidence of a relationship between precision and underlying acuity at the 5% level (PC10-test: t=-0.69, p=0.50; PC5-test: t=-0.33, p=0.75; ETDRS1: t=1.09, p=0.29; ETDRS2: t=0.54, p=0.60). However, an unexpected finding related to the spread of test-retest points for the ETDRS charts when plotted separately for the two versions of the ETDRS chart. Inspection of Figs 8.3.5 and 8.3.6 suggests a considerably wider spread of test-retest points for version 1 of the ETDRS chart compared with version 2. This would be expected to translate to a difference in precision between the two charts. To examine this further, the 95% TRR was estimated for each version of the chart separately. This resulted in a 95% TRR of ±0.23 logMAR for the ETDRS-1 chart, and ±0.13 logMAR for ETDRS-2. This difference in the width of the 95% TRR equates to 1 ETDRS line, and reaches statistical significance at the 5% level (F=0.333, p=0.025).

Figure 8.3.3. Bland-Altman plot for precision: PC10-test

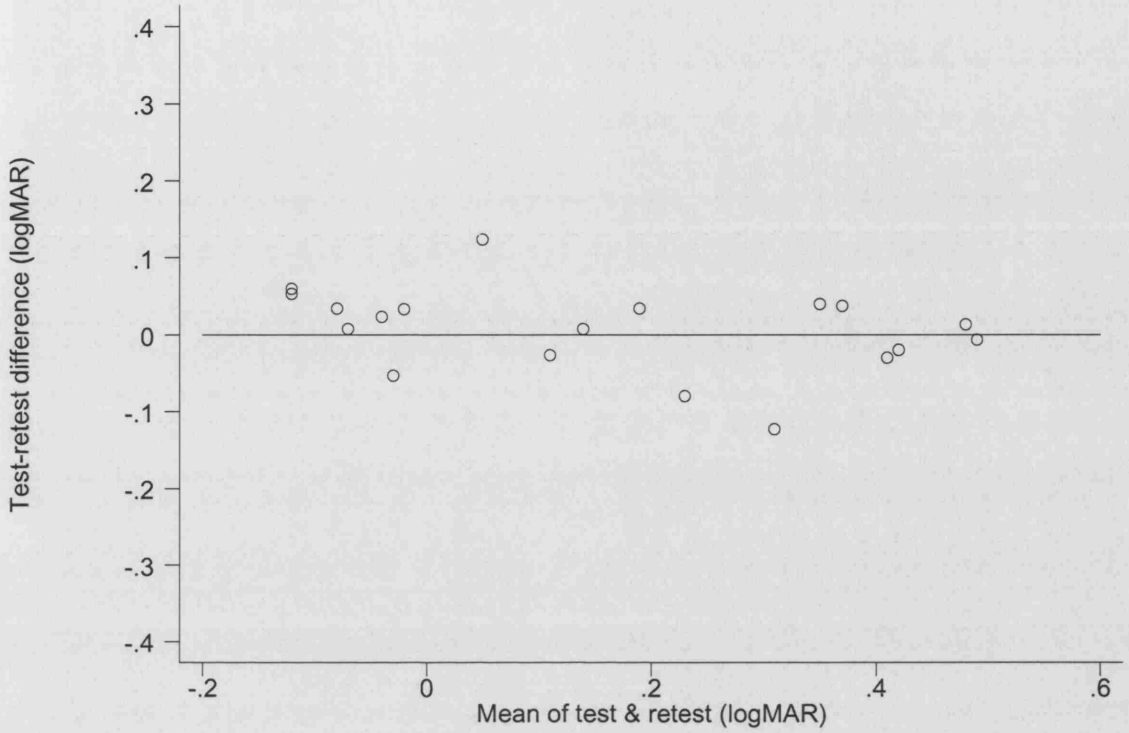


Figure 8.3.4. Bland-Altman plot for precision: PC5-test

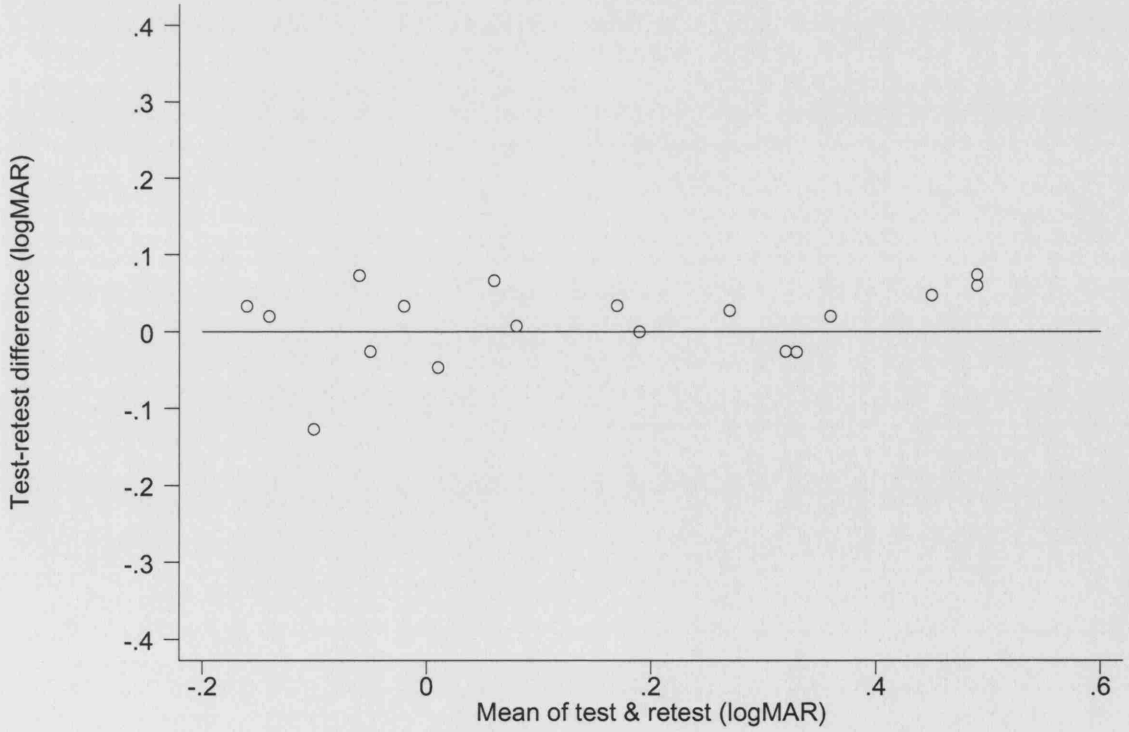


Figure 8.3.5. Bland-Altman plot for precision: ETDRS 1

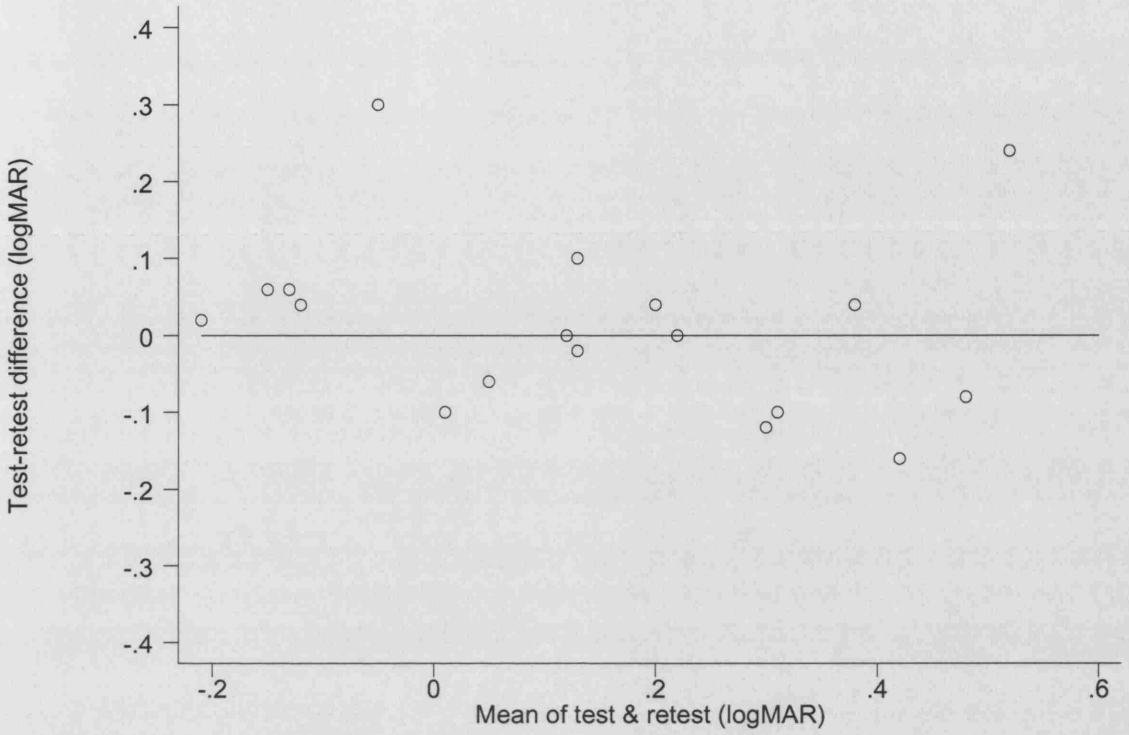
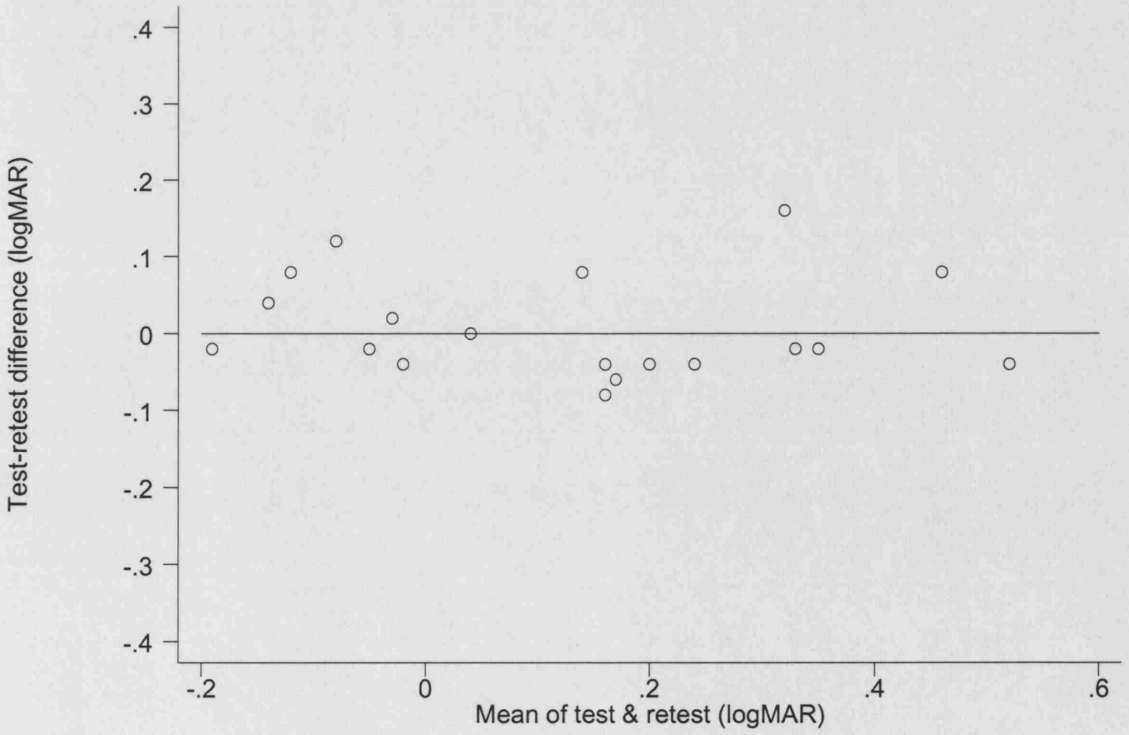


Figure 8.3.6. Bland-Altman plot for precision: ETDRS 2



Measurement Time

Table 8.3.5 shows the measurement times for both the PC10- and PC5-tests. It can be seen that the median measurement time for the PC5-test is almost half that of the PC10-test. This represents an average time saving of nearly 4 minutes per measurement.

Table 8.3.5. PC-test measurement time

Test	Time (secs)		
	5^{th} centile	50^{th} centile	95^{th} centile
PC10-test	274	518	735
PC5-test	166	284	520
ETDRS*	50	109	274

*data from earlier study (section 5.3)

The duration of an ETDRS measurement was not formally assessed in this study. However, the data from an earlier study (section 5.3) has been included for comparison. The median ETDRS measurement time is approximately 40% of that for the PC5-test.

9. DISUSSION

The results of this study suggest that it is possible to measure visual acuity using a computerised test of the type described in section 7.2. Acuities measured using both the PC10- and PC5-tests were, on average, slightly worse than those measured using the ETDRS chart (see Table 8.3.2). However, the degree of bias (equivalent to between 1 and 1½ ETDRS letters) did not reach statistical significance at the 5% level, and was not was not related to the underlying level of acuity. Possible reasons for the small degree of bias between the two tests are listed in Table 9.1.

Table 9.1. Comparing the PC-test and ETDRS tests: possible sources of bias

1) Less contour interaction in the PC-test[A]
2) Pixelisation effects
3) Lower test luminance of the PC-test

If differences in the degree of contour interaction had exerted an influence, it would have been expected to create bias in the opposite direction than was observed in this study. Also, if present, a pixelisation effect would have been more influential for better acuities for which pixel size is relatively large compared with stimulus size. The lack of any acuity-related bias therefore suggests that pixelisation effects did not influence the results. The difference in test luminance is also a potential cause of the observed degree of bias. The relationship between luminance and visual acuity over the luminance range featured in this study is not strong (see section 2.10.9). However, the effect of increased luminance may have been exaggerated by reducing pupil size and therefore retinal image blur in those with residual refractive errors. It is also, conceivable that any detrimental effect of lower luminance on acuities measured with the PC-test may have been mitigated to an extent by reduced contour interaction.

[A] 3 of the 5 letters on each line of the ETDRS test are surrounded by contours on both sides, whereas the use of only 3 letters per line in the PC-test results in only 1 out of three having contours on both sides.

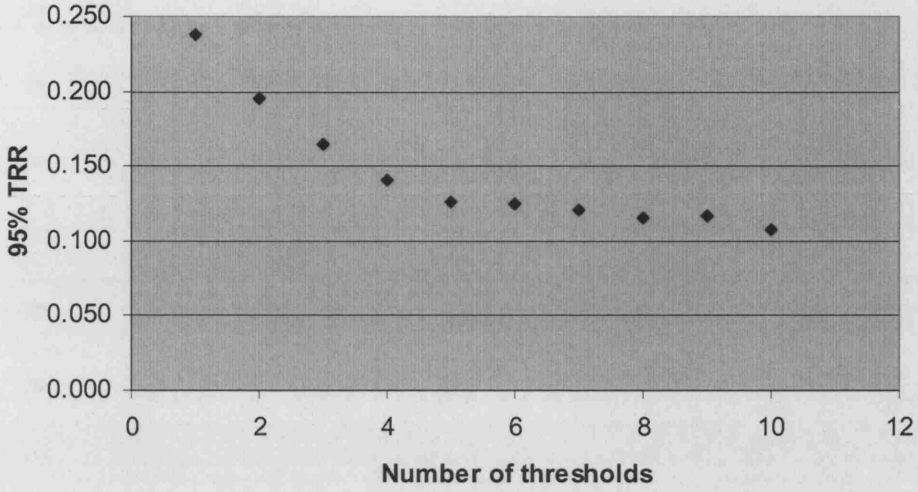
The precision of the PC10-test and PC5-test in terms of the 95% TRR was ±0.11 logMAR and ±0.10 logMAR respectively (see Table 8.3.4). The corresponding value for the pooled data from ETDRS charts 1 and 2 was ±0.18 logMAR. This estimate represents a significantly poorer degree of precision than was observed with either the PC10-test (p=0.014), or the PC5-test (p=0.005). This difference in precision translates into a considerable delay in detecting genuine changes in acuity (see section 2.9). Based upon this estimate of precision for the ETDRS chart, a subject's acuity measurement would have to change by ±0.20 logMAR before the change could be attributed to a true alteration in clinical status rather than measurement error alone. This corresponds to a change in smallest resolvable letter size of approximately 58%. The corresponding change for the PC5-test would be ±0.11 logMAR which is equivalent to a 29% change in smallest resolvable letter size. Based upon these results, the increased precision of the PC5-test compared with the ETDRS chart represents a significant improvement in the ability to detect true clinical change.

The estimate of precision for the PC10-test in this study (±0.11 logMAR) was slightly worse than that obtained for the PC5-test (±0.10 logMAR). This difference is equivalent to ±½ an ETDRS letter (±0.01 logMAR) and was not statistically significant at the 5% level (F=1.196, p=0.713). This result suggests that the approach of repeating and averaging acuity thresholds offers no additional benefit over and above 5 repeated thresholds for a test of this design.

In an attempt to gain further information as to the optimum number of repeated thresholds, a technique similar to that utilised in the pilot study described in section 7.3.3 was applied to the paired test-retest measurements for the PC10-test. The acuity scores for all 19 subjects who underwent a PC10 test on both visits were recalculated taking only the first of the 10 repeated thresholds into account. This was then repeated for the first two thresholds, then the first three, and the process was continued up to the total of 10 thresholds. This allows us to recalculate the 95% TRR for

each number of thresholds from 1 up to 10. Fig 9.1 shows the width of the 95% TRR plotted against the number of thresholds contained within a PC-test. The width of the 95% TRR is seen to reduce dramatically as the number of thresholds increases from 1 to 5. Beyond 5 repeated thresholds, the improvement is slight. The 95% TRR corresponding to the first 5 thresholds only of the PC10 test is ±0.13 logMAR. This is consistent with the difference in the width of 95% TRR estimated for the PC10-test in this study (±0.11 logMAR) as compared with the PC5-test (±0.10 logMAR) being due to chance alone. The appearance of Fig 9.1 is very similar to that of Fig 7.3.1 which relates to a similar analytical approach utilised in the pilot study (section 7.3.3). This suggests that the use of the standard error of the mean in the pilot study as a predictor of the relationship between number of acuity thresholds within a PC-test measurement and the test's precision was appropriate.

Figure 9.1. The relationship between number of PC-test thresholds and precision (in terms of the 95% TRR)



The width of the 95% TRR for the ETDRS chart as calculated from pooled data from charts 1 and 2, was ±0.18 logMAR. This was identical to the estimate for the ETDRS chart in the initial study (see sections 5 and 6, and Table 5.3.5). However, when Bland-Altman plots were generated for the two charts separately (Figs 8.3.5-6) it was noted that the vertical spread of points appeared to be larger for ETDRS chart 1 as compared with chart 2. Estimates of precision for the individual charts were consistent with this observation. The 95% TRR for chart 1 (±0.23 logMAR) being significantly wider (F= 0.333, p=0.025) than that for chart 2 (±0.13 logMAR). Such a difference in performance was surprising in view of the fact that the two charts are designed to be equivalent. Possible explanations for this discrepancy include:

1) a genuine difference exists between the performance of the two charts,
2) experimental error has occurred, or
3) this observation has occurred by chance.

The first of the above explanations seems unlikely in the light of the rigorous principles underlying the design of the ETDRS chart. The charts differ only in terms of the letter combinations used, and these are carefully

selected such that the average difficulty of each line is equivalent (section 2.13). Experimental error as a cause of this discrepancy would most probably take the form of a large outlier. Examination of Fig 8.3.5 is not suggestive of a single outlying point. This, along with the lack of evidence of non-normality for the distributions of test-retest differences (Table 8.3.3) does not support experimental error as the cause of the discrepancy. As a precaution, all acuity scores for both versions of the ETDRS chart were recalculated from the data proformas on which subject responses were recorded. No errors were identified for either chart. Accordingly, a chance finding was considered the most plausible explanation. Although the number of subjects used in this study exceeds that used in many other similar published experiments, the use of a relatively small number of subjects will increase the chance of such an occurrence. Assuming the observed difference between the two estimates of ETDRS precision to be due to chance, we might expect their mean to better reflect the true level of precision. With this in mind, it is interesting to note that the mean of the precision estimates for the two ETDRS charts (±0.18 logMAR) is identical to the estimate of ETDRS precision obtained in the study described in section 8. It is also possible, based upon a knowledge of the differences in test design to predict the precision of the ETDRS chart from the estimate of precision obtained for the PC5-test. In theory, the PC5-test is equivalent to an ETDRS test with 15 letters per line (see section 2.10.6). Therefore, if the precision of a 15 letter-per-line ETDRS test is ±0.10 logMAR, then the predicted precision of a 5 letter-per-line ETDRS test would be larger by a factor of $\sqrt{3}$ (as the scale of the 15 letter-per-line chart is 3 times finer). This results in a predicted ETDRS precision of ±0.17 logMAR, which is close to the average of the two estimates obtained in this study. This appears to support the conclusion that the observed difference in the 95% TRR for ETDRS charts 1 and 2 was due to a chance finding.

In the absence of an opportunity to recruit further subjects, it was noted that a further comparison of the performance of the ETDRS chart and PC5-

test on a larger number of subjects may be obtained through a further analysis of the data from the first visit only. On visit one of the study, acuity measurements were taken using four tests in total: versions 1 and 2 of the ETDRS chart, along with the PC10-test and the PC5-test. Recalculating the PC10-test scores taking account of only the first 5 of the test's 10 thresholds results in a measurement which is equivalent to one taken with a PC5-test. This produces two series of paired measurements, one for the ETDRS chart (with one measurement taken using each version) and one for the PC5-test (with one measurement being a PC5-test measurement and the other being the first half of a PC10-test). These two series of measurements were analysed for 34 subjects of the 35 who attended the first visit[A], resulting in an estimated 95% TRR of ±0.13 logMAR for the ETDRS chart, and ±0.11 logMAR for the PC5-test[B] (see Figs 9.2 and 9.3). Comparing the results of this additional analysis with the initial set of results provides some interesting observations. Firstly, the estimate of PC5-test precision obtained from paired measurements performed on one visit is identical to that obtained from paired measurements separated by at least two weeks. This observation suggests that estimates of precision are relatively independent of the length of time between measurements. Assuming this to be the case, we can compare the estimate of ETDRS precision derived from 34 subjects over 1 visit (±0.13 logMAR) with the two estimates of ETDRS precision derived from 19 subjects over 2 visits (ETDRS-1: ±0.23 logMAR; ETDRS-2: ±0.13 logMAR). In the light of evidence that estimates of precision from paired measures are independent of the intervening time, these results are perhaps more consistent with ±0.13 logMAR as being representative of the precision of the ETDRS chart than the figure of ±0.23 logMAR.

[A] Excluding subject 17 for whom there was a missing observation (see section 8.3 paragraph 1).
[B] These estimates of precision were not statistically different at the 5% level (F=0.79, p=0.50).

Figure 9.2. Precision for ETDRS chart (visit 1 only, 34 subjects)

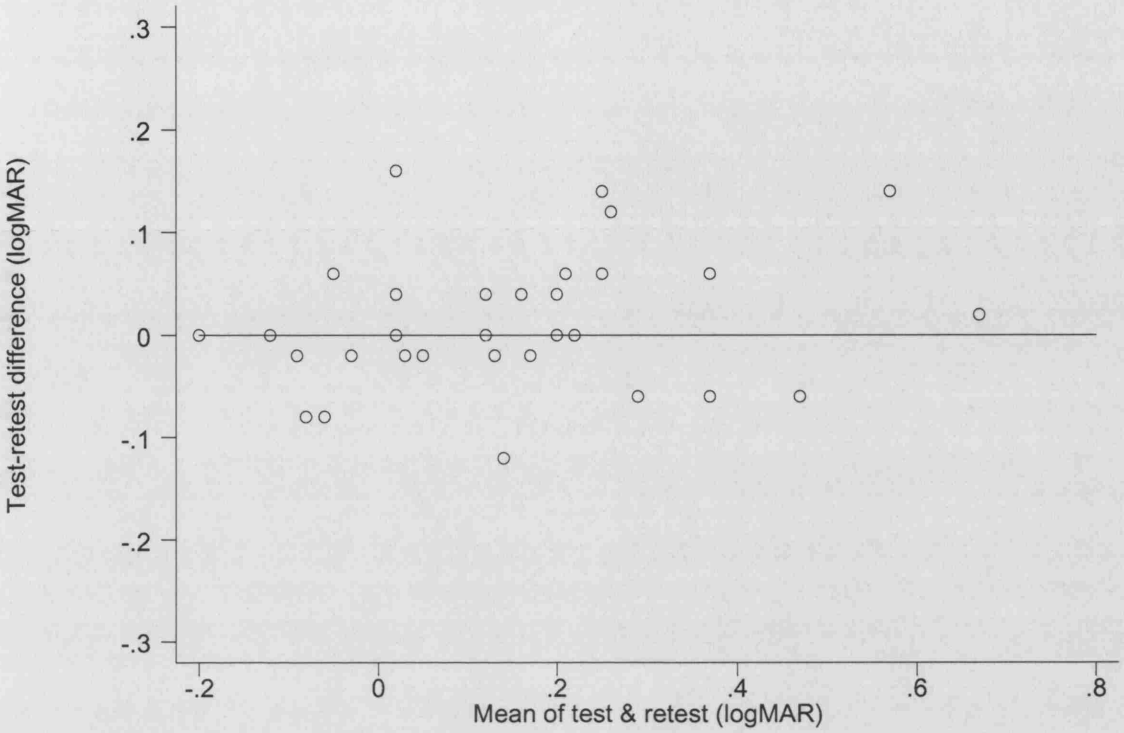
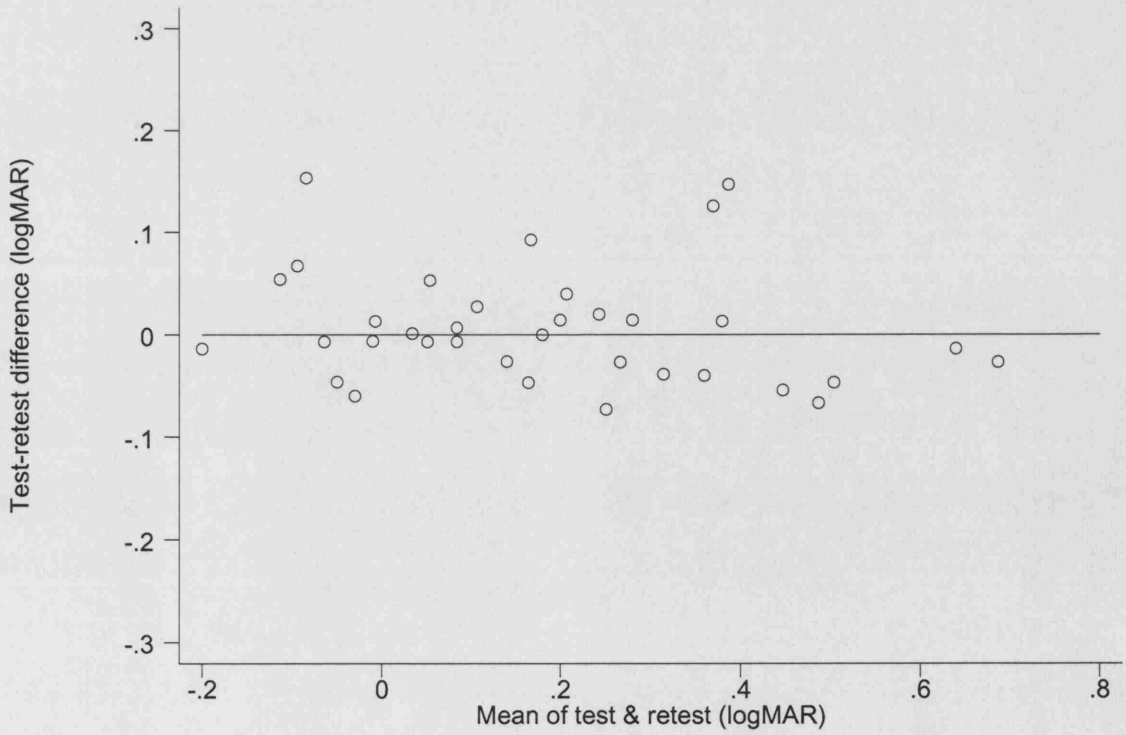


Figure 9.3. Precision for PC5-test (visit 1 only, 34 subjects)



In summary, the results of this study suggest that visual acuities measured with a computerised test as described herein are accurate when compared with those of the ETDRS chart. The results also suggest that the optimum number of thresholds for a computerised test of this design is 5, a larger number of thresholds offering no further improvement in precision. The initial analysis suggested that the PC-test acuities were more precise than ETDRS acuities. However, the discrepancy between estimates of precision for the two versions of the ETDRS test, along with the results of a additional analysis on a larger subject group, suggests that random error may have resulted in an underestimate of precision for the ETDRS chart. A genuine difference in the precision of the two ETDRS chart versions cannot be ruled out, but appears unlikely in view of the design principles on which the charts are based. In view of the equivocal nature of the estimate of ETDRS precision obtained in this study, it would be desirable to obtain further estimates of precision for both tests using a larger number of subjects (see section 14).

Link section

An interesting finding in the previous study was the observation that the differences in test design between the computerised test and the ETDRS chart did not appear to induce any notable bias between the acuities measured using the two tests (Table 8.3.2). The differences between the two tests which were thought to be potential sources of bias are listed in Table 9.1. Examination of the data appeared to discount pixelisation effects as a source of bias (see section 9). The conclusion was therefore drawn that either differences in contour interaction and luminance were not influential, or that their opposing effects cancelled one another out. As the influence of contour-interaction is known to be greater in amblyopic subjects, additional information as to the influence of contour-interaction may be gained by comparing the results of the two tests in an amblyopic population. Such a comparison would also provide an opportunity to assess test performance over a wider range of acuity, as well as assessing whether test precision was related to the underlying level of acuity.

10. COMPARING THE PERFORMANCE OF THE PC5-TEST AND ETDRS CHART IN AMBYLOPIC SUBJECTS

10.1. AIM

- To assess the accuracy of the PC-test compared with the ETDRS chart in amblyopic subjects.
- To compare the precision of the PC-test with that of the ETDRS chart in amblyopic subjects.

10.2. METHODS

10.2.1. *Subjects*

Subjects were recruited from amongst those attending for routine orthoptic outpatient clinic appointments at Moorfields Eye Hospital NHS Trust, London, UK. Inclusion criteria were as follows:

- A diagnosis of strabismic amblyopia in one eye
- An inter-eye difference in acuity of two or more Snellen lines
- Visual acuity (with habitual spectacle correction if worn) of Snellen 3/60 or better
- 14 years of age or over
- Able to understand and comply with the testing protocol

10.2.2. *Equipment*

ETDRS logMAR chart

The ETDRS chart is described in full in section 2.13. The charts were back-lit in the standard Lighthouse box achieving a luminance of 300 cd/m^2 and contrast of 98%. Versions 1 and 2 of the chart were used.

PC5-test

The PC-test is described in section 7.2. The version used in this study was that featuring 5 acuity thresholds per single PC-test measurement. The details of the tests are summarised in Table 10.2.1. A 30 minute warm-up period was employed prior to collection of any data to allow screen luminance to stabilise. Following this period luminance was measured at 90 cd/m^2 with a contrast of 92%.

Table 10.2.1. Summary of the tests under comparison

Test	Letters per line	Line Increment (logMAR)	Thresholds*	Jump*	Scale increment** (logMAR)
ETDRS	5	0.10	1	N/A	0.020
PC5-test	3	0.10	5	5	0.007

* - see section 7.2.3 for definitions

** - Scale increment = $\frac{\text{(Inter-line increment)}}{\text{(No of letters per line) x (No of thresholds)}}$

10.2.3. *Scoring*

ETDRS chart

The ETDRS chart was scored using the interpolated scoring method (see section 4.2.6). An endpoint of a full row of errors was used (see section 4.2.6).

PC5-test

The algorithm for this test is described in detail in section 7.2. The algorithm employs an identical scoring method, and termination rule to that used with the ETDRS chart.

10.2.4. *Investigations*

Four acuity measurements were taken on one eye of each subject. Two ETDRS acuity measurements were taken, one using version 1 and one using version 2. Two measurements were also taken using the PC5-test. The testing order was randomised (see section 4.2.7). All subjects wore their habitual spectacle correction. Due to the wider range of acuities included in this study, the provision for measuring acuities from two different testing distances was required. Prior to commencing the four randomised measurements, subjects were asked to attempt 3 stimuli at the +0.80 logMAR on the PC5-test (the starting stimulus size for the PC5-test) from a viewing distance of 4 metres. If all of these stimuli were correctly named, all subsequent measurements were carried out using a viewing distance of 4 metres. If any of these 3 stimuli were misnamed, all subsequent measurements were carried out using a 1 metre viewing distance.

10.3. RESULTS

22 subjects underwent testing using the regimen described above. Age at last birthday ranged from 16 to 52 years with a median of 35 years. Acuity (as measured with version 1 of the ETDRS chart) ranged from +0.02 to +1.24 logMAR (Snellen equivalent 6/6.3 to 6/104) with a median of +0.60 logMAR (Snellen equivalent 6/24). This required 13 measurements to be carried out at a testing distance of 4 metres, and 9 using a 1 metre testing distance.

Accuracy

The accuracy of the PC5-test as compared with the reference standard ETDRS chart was assessed using the methods of Bland and Altman (section 4.1). Table 10.3.1 shows the mean difference between a PC5-test measurement and an ETDRS measurement. The mean difference is slightly less than zero showing that, on average PC5-test acuities were slightly better than those measured with the ETDRS chart. The magnitude of this bias equated to approximately 1½ ETDRS letters. The 95% confidence interval for the mean bordered on zero suggesting that this degree of bias was close to reaching statistical significance.

Table 10.3.1. Accuracy as compared with reference standard ETDRS chart

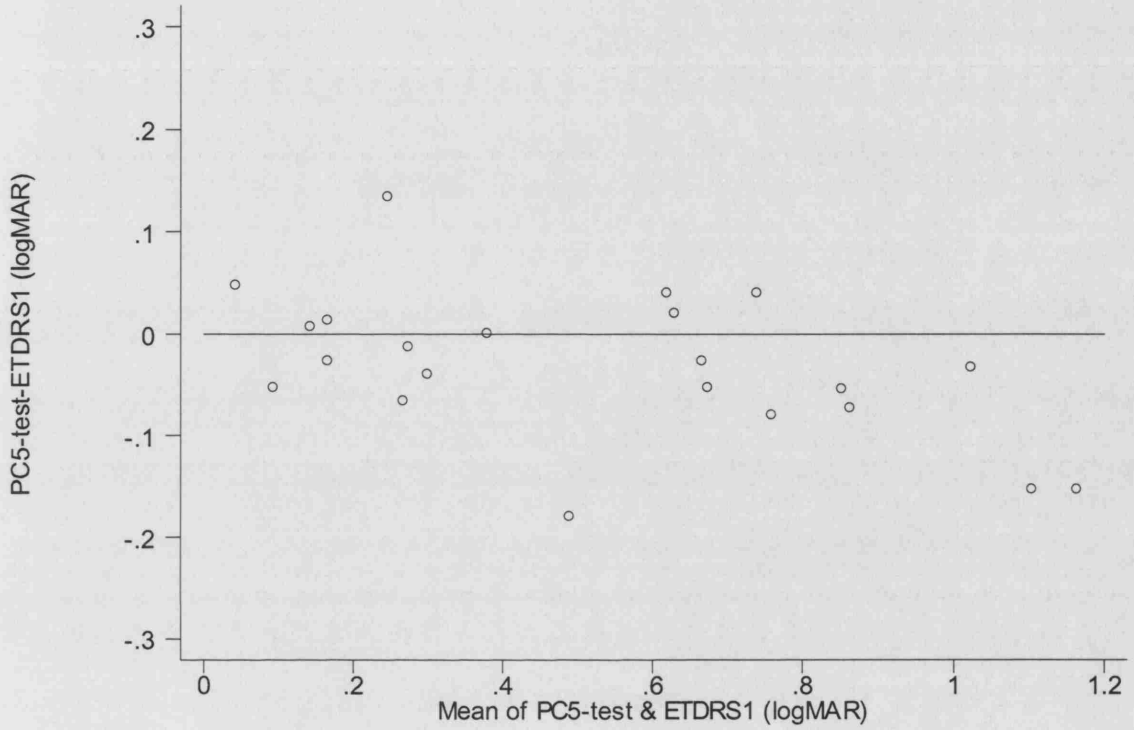
Tests	Mean difference (logMAR)	95% CI (logMAR)
PC5-test – ETDRS 1	-0.032	-0.064 to 0.00 logMAR

Acuity-related bias

Fig 10.3.1 shows a Bland-Altman plot in which the difference between PC5-test and ETDRS measurements is plotted against their mean. The points towards the left of the plot are situated evenly above and below the midline, whereas those towards the right are predominantly below the midline. This suggests that for mild degrees of amblyopia, the PC-test is accurate compared with the ETDRS chart, whereas for greater degrees of amblyopia the PC-test tends to produce better acuities than the ETDRS chart. Linear regression analysis (see section 4.1.6) was suggestive of a

statistically significant relationship between the difference between the two tests and their mean (t=-2.53, p=0.02).

Figure 10.3.1. Bland-Altman plot: PC5-test vs ETDRS



Precision

The distribution of test-retest differences for both charts were assessed for evidence of departure from a normal distribution using the Shapiro-Wilk W-test (see section 4.1.5). There was no evidence of such a departure for either the PC5-test (W=0.968, p=0.671) or the ETDRS chart (W=0.942, p=0.215). The methods of Bland and Altman were therefore used to assess test precision (section 4.1.2). Table 10.3.2 shows the estimates of precision for both tests. The width of the 95% TRR for the ETDRS chart was ±0.05 logMAR wider than for the PC5-test. This difference equates to ½ a line of ETDRS letters, and approached statistical significance at the 5% level.

Table 10.3.2. Precision of the PC5-test as compared with the ETDRS chart

Tests	95% TRR (logMAR)	F* (cf ETDRS)
PC5-test	±0.10	0.449 (p=0.074)
ETDRS	±0.15	-

* - F-test (see section 4.1.6)

Figs 10.3.2 and 10.3.3 show Bland-Altman plots for precision for the PC5-test and the ETDRS chart respectively. The reduced precision of the ETDRS chart compared with the PC5-test is indicated by the wider spread of differences between test and re-test. Neither plot exhibits a wider vertical spread of points towards the right side of the plot suggesting that there was no relationship between precision and underlying acuity. This was confirmed by linear regression of the unsigned differences against the mean of the two test results (PC5-test: t=1.57, p=0.133; ETDRS: t=0.20, p=0.843).

Figure 10.3.2. Bland-Altman plot: PC5-test

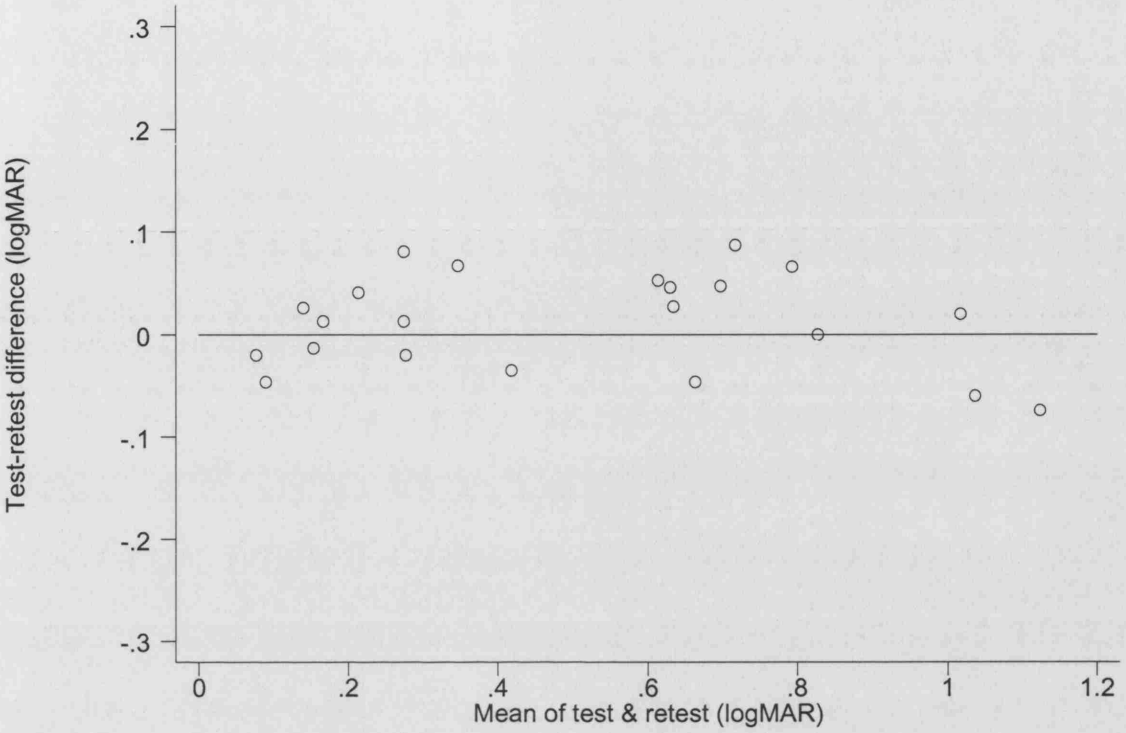
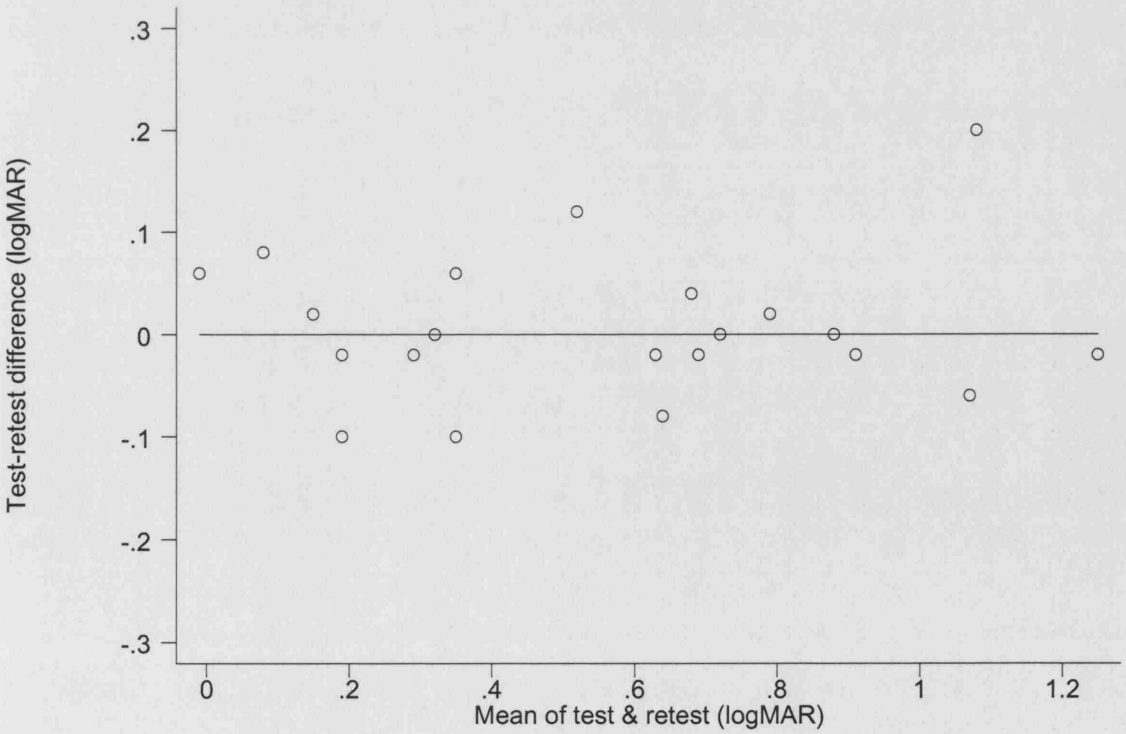


Figure 10.3.3. Bland-Altman plot: ETDRS



11. DISUSSION

In this study the PC5-test tended to produce slightly better acuities as compared with the ETDRS chart. This magnitude of this bias was relatively small being approximately equivalent to 1½ ETDRS letters (Table 10.3.1). In both this study and the previous study (described in section 8 and discussed in section 9), the degree of bias did not reach statistical significance at the 5% level (see Tables 10.3.1 and 8.3.1). Despite this, an interesting observation was that the direction of bias was opposite in the two studies. In non-amblyopes (the previous study- section 8) PC-test acuities were on average ¼ ETDRS line worse than ETDRS acuities, whereas in amblyopic subjects (the present study – section 10), PC-test acuities were approximately ¼ ETDRS line better. A possible explanation for this relates to two of the differences in test design listed in Table 9.1: the lower luminance of the PC-test, and the fact that fewer letters are subject to contour interaction from both sides compared with the ETDRS chart. These two potential sources of bias will tend to oppose one another but whereas the effect of luminance is likely to be similar in amblyopes and non-amblyopes, the influence of reduced contour interaction would be expected to be greater amblyopes. Hence in the present study, the effect of reduced contour interaction may have more than compensated for the lower luminance level producing bias in the opposite direction to that seen in the previous study.

This conclusion would appear to be supported by the degree of acuity-related bias observed in the present study. Fig 10.3.1 suggests that for mild degrees of amblyopia, PC-test acuities are accurate compared with ETDRS acuities. However, for greater degrees of amblyopia, the PC-test tends to produce better acuities than the ETDRS chart. In view of the knowledge that the effect of contour interaction upon acuity in amblyopia is greater for higher degrees of amblyopia (see section 2.10.8), this finding appears consistent with variable contour interaction as the explanation for the opposing degrees of bias in the two studies.

The precision (in terms of the 95% TRR) of the PC5-test and ETDRS chart in this study was ±0.10 logMAR and ±0.15 logMAR respectively. This difference equates to ±½ an ETDRS line, and approached significance at the 5% level (p=0.07). The precision of the PC5-test in this study is identical to that obtained in the previous study (described in section 8 and discussed in section 9), suggesting that the precision of the PC5-test is unlikely to be much worse for poorer acuities, at least where the cause is amblyopia. This conclusion would appear to be supported by the lack of a relationship between precision and underlying acuity in this study. The estimate of precision for the ETDRS chart in this study (±0.15 logMAR) was slightly better than the average of the two estimates in the previous study (±0.18 logMAR). Although there was an unexpected discrepancy between the two estimates of ETDRS precision obtained in the previous study, it appears unlikely that the precision of the ETDRS chart is very different for amblyopic eyes than for non-amblyopic eyes.

In summary, this study has shown that the PC5-test demonstrates a good level of accuracy when compared with the ETDRS chart in amblyopic subjects. The precision of the PC5-test was ±0.05 logMAR (½ an ETDRS line) better than that of the ETDRS chart. Although the magnitude of this difference may be considered significant from a clinical perspective, with only 22 subjects in the study, the difference fell short of reaching statistical significance at the 5% level (p=0.07). Although not conclusive, combining the results of this study with that of the previous one allows us to speculate further as to the relationship between the design of a visual acuity test and its performance[A]. For example, it appears likely that the reduced degree of contour-interaction exhibited by the PC5-test causes an underestimate of the visual deficit in amblyopia of around 0.05 logMAR. However, because the effect of contour-interaction is less in non-amblyopic subjects, the influence of the differing degrees of contour-interaction is likely to be minimal in non-amblyopic subjects. The results of these studies

[A] Any conclusions drawn from a comparison of the results of these two studies are strengthened due to the fact that both studies were carried out in the same room by the same examiner using identical equipment.

may form the basis for modifying the design of the PC5-test to improve its accuracy in the presence of varying degrees of sensitivity to contour-interaction.

Link section

The results of the studies described in sections 8 and 10, and discussed in sections 9 and 11 respectively, suggest that the average of a series of multiple thresholds of acuity may be more precise than a single estimate of acuity measured using the ETDRS chart. However, estimates of ETDRS precision within the first study were inconsistent, and a larger study is required to confirm this finding. A considerable degree of variability exists between published levels of visual acuity test precision, with estimates of the width of the 95% TRR ranging from ±0.07 logMAR[74 78] to ±0.26 logMAR[81]. Some of this variation may be explained by differences in test design and/or scoring method. However, even if we consider only those published estimates made using charts based upon the Bailey-Lovie design (section 2.13) and which feature interpolated scoring, the range of reported values, although reduced, is still substantial (2.7 fold variation, Table 11.1).

Table 11.1. Published estimates of precision using charts based on the Bailey-Lovie design.

Author	95% TRR	No of test sessions	Subjects	Refractive correction
Elliott et al[74]	±0.07	2	Normals	Full
van den Brom et al[211]	±0.08	1	Cataract	Full
Elliott et al[74]	±0.09	2	Cataract	Full
Arditi et al[79]	±0.09	2	Normals	Full
Bailey et al[80]	±0.10	1	Normals	Full
Rosser[Δ]	±0.15	1	Amblyopes	Habitual
Lovie-Kitchin[81]	±0.16	1	Normals	Unaided
Rosser[ΔΔ]	±0.18	1	Cataract	Habitual
Reeves et al[84]	±0.19	2	Mixed	Full

[Δ] – indicates the estimated 95% TRR from the study described in sections 8 and 9
[ΔΔ] – indicates the estimated 95% TRR from the study described in sections 5 and 6

What are the causes of this large variation? No association between age and precision was found in the studies of Beck et al[193] and Lovie-Kitchin et

al[144]*. Reeves*[83] *suggested three other potential sources of variation in precision: the presence or absence of disease, whether or not the measurements were conducted in a single session, and the presence or absence of uncorrected refractive error. Subjects with cataract feature at both extremes of the 95% TRR in Table 11.1, suggesting that presence of cataract is not a major determinant of precision. The combined results of the previous two studies described in this thesis (sections 8 to 11) suggest that the presence of amblyopia is unlikely to be associated with reduced precision. Three of the four studies from Table 11.1 which have estimated precision over two visits reported 95% TRRs of less than ±0.10 logMAR, which is not consistent with a hypothesis of reduced precision with longer periods between examinations. Section 8 of this thesis presented data which are also not consistent with such a hypothesis*[A]*.*

Although there has been speculation that a relationship may exist between uncorrected refractive error/optical defocus and the precision of visual acuity measurements[83]*, the empirical data are ambiguous. Siderov et al*[212] *reported a 95% TRR of ±0.16 logMAR for both aided and unaided measurements (although the study methodology appears to lack consistency in terms of the design of chart used for test and retest). By contrast Elliott et al*[74] *found the width of the 95% TRR in unaided normal subjects to be 3 times that observed when a full correction was worn (95% TRR of ±0.21 logMAR and ±0.07 logMAR respectively). Neither publication details the amount of refractive error in the study population. The only published study which has systematically investigated the effects of optical defocus in a way that sheds light upon its influence on precision, is that of Carkeet*[213]*. This study assessed the effect of two discrete levels of optical defocus on the shape of the frequency-of-seeing curve as determined by Probit analysis*[214]*. The results showed that optical defocus was associated with a flattening of the frequency-of-seeing curve. Based on this finding,*

[A] The study described in section 8 showed that the precision of a computerised visual acuity estimated from two measurements taken in the same session was identical to that derived from two measurements separated by 2 weeks.

Carkeet commented that the 95% TRR would be expected to be wider under conditions of defocus than for well corrected subjects.

12. TO WHAT EXTENT DOES UNCORRECTED REFRACTIVE ERROR INFLUENCE THE PRECISION OF VISUAL ACUITY MEASUREMENTS?

12.1. AIM

- To investigate whether a relationship between optical defocus and the precision of visual acuity measurements exists in normal subjects.

12.2. METHODS

12.2.1. *Subjects*

Normal subjects were recruited from the staff of Moorfields Eye Hospital NHS Trust. Inclusion criteria were as follows:

a) Age 50 years or less,

b) Refractive error (mean sphere) within the range +0.50D to -10.00D,

c) Astigmatism not exceeding 1.50D,

d) Absence of any ocular abnormality including media opacity,

e) No history of ocular abnormality including amblyopia,

f) No history of regular use of the ETDRS logMAR chart, and

g) Acuity better than +0.20 logMAR (Snellen equivalent 6/9.5).

12.2.2. *Testing procedure*

Prior to testing, each subject underwent formal subjective refraction[A] as described in section 4.2.4. The chart used for refraction was used for that purpose only. One eye only of each subject was assessed (section 4.2.5). Where both eyes met the inclusion criteria, the right eye was used as the study eye. Each subject then underwent 2 acuity measurements with each of three refractive corrections:

1. Full refractive correction (zero optical defocus),
2. Full refractive correction plus +0.50D defocus, and
3. Full refractive correction plus +1.00D defocus.

Each subject therefore underwent a total of 6 acuity measurements using the ETDRS logMAR acuity chart (section 2.13) in a session lasting approximately 15 to 20 minutes. All three versions of the ETDRS logMAR chart were used; the refraction chart, chart 1, and chart 2. These were used for the measurements at zero defocus, +0.50D defocus, and +1.00D defocus respectively. The refraction chart was used for the measurements at zero defocus because the control of letter difficulty is less rigorous than for charts 1 and 2 (section 2.13). Accordingly, should the refraction chart be associated with a lower degree of precision than charts 1 and 2, this would lessen

[A] The starting point for subjective refraction was determined by objective refraction by retinoscopy. Any subject with an objective refraction >+0.50D was considered to have latent hyperopia and excluded.

any measured effect of defocus on precision rather than exacerbate it. To limit the influence of memory effects when using 3 charts for 6 measurements, subjects were required to read the chart forwards for one measurement, and backwards for the other. The order of the six measurements was randomised with the sole restriction of the same chart never being used for consecutive measurements (see section 4.2.7). All charts were viewed from a distance of 6.3 metres. This results in a range of effective letter sizes from +0.80 to -0.50 logMAR (Snellen equivalent 6/37.9 to 6/1.9) as opposed to the range at the standard 4 metre distance of +1.00 to -0.30 logMAR (Snellen equivalent 6/60 to 6/3). The 6.3 metre testing distance was chosen to prevent subjects with very good acuity reading all or part of the bottom line on the chart. It is desirable to prevent this as the resultant 'truncation' or 'ceiling' effects may artificially reduce the TRV for those with good acuity[83]. A longer testing distance was avoided because, for a testing distance of 6.3 metres, there would be three rows of letters above the +0.50 logMAR letter size which is the expected threshold for 1 dioptre of optical defocus[215 216]. The subject's responses were recorded on specially designed data proformas (see section 4.2.10).

12.2.3. *Scoring and endpoints*

Scoring and termination rules were chosen to maximise test precision. Each chart was attempted using a forced-choice testing paradigm (see section 4.2.8). The end point for each measurement was taken as a full row of errors (see section 4.2.8). Each chart was scored using the interpolated scoring method to maximise precision (see sections 2.10.6 and 4.2.9). For each subject, at each level of optical defocus, the difference between the two acuity measurements was calculated by subtracting the score from the backwards measurement from that from the forwards measurement.

12.2.4. *Statistical methods*

For each individual the difference in their two acuity measurements at a given degree of defocus was calculated. The standard deviation of these differences across all 40 individuals was estimated. Precision was quantified in terms of the 95% TRR (see section 4.1.2). Evidence of differences in standard deviations across the three degrees of defocus was sought using Levene's test (see section 4.1.6) which is robust to

departures from normality. The assumption of normality itself was assessed for each degree of defocus (zero, 0.50D and 1.00D) using the Shapiro-Wilk W-test and by inspection of quantile-normal plots (see section 4.1.5).

12.3. RESULTS

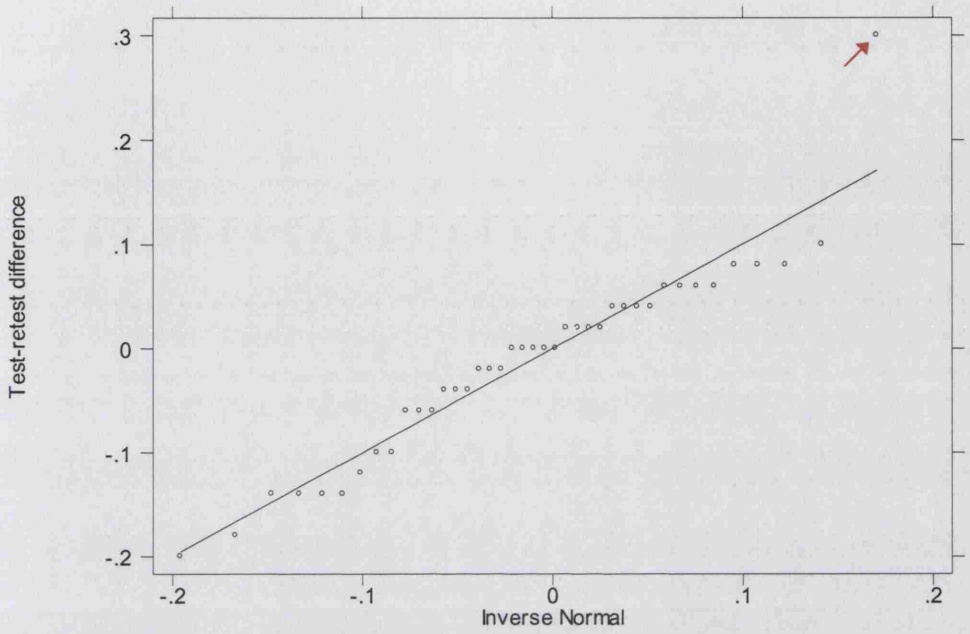
40 subjects were recruited. Age at last birthday ranged from 21 years to 50 years (median 33 years). Refractive error ranged from +0.50D to -8.75D (median spherical equivalent -1.25D). Astigmatism ranged from zero to 1.25D (median zero). Acuity ranged from -0.32 to +0.12 logMAR with a median of -0.14 logMAR (Snellen equivalents: range 6/3 to 6/8, median 6/4.3). Three individuals had acuity worse than 0.00 logMAR (Snellen 6/6), the worst of which was +0.12 logMAR (Snellen 6/8).

Table 12.3.1. Mean differences between paired acuity measurements and their 95% ranges by degree of optical defocus

Degree of optical defocus	Mean difference (95% CI) (logMAR)	sd (logMAR)	95% TRR (logMAR)
0.00D	+0.002 (-0.014 to +0.018)	0.051	±0.10
+0.50D	-0.013 (-0.043 to +0.017)	0.093	±0.18
+1.00D	-0.004 (-0.044 to +0.036)	0.124	±0.24

The mean difference at each degree of defocus was small and the 95% confidence interval included zero (Table 12.3.1) indicating that, in this group of subjects, the ETDRS charts were no more difficult when attempted backwards. There was strong evidence that the standard deviation of the differences varied with the degree of defocus ($p=0.0002$), with the standard deviation increasing (and hence precision decreasing) as the level of optical defocus increased. There was no evidence of non-normality for the distribution of differences for either zero defocus or defocus of +1.00D ($p=0.38$). At +0.50D defocus there was weak evidence against the hypothesis of normality ($p=0.04$). This was largely due to one observation with an outlying value of +0.30 logMAR (indicated by a red arrow in Fig 12.3.1). Excluding this observation from the analysis resulted in a 95% TRR based on normal theory of ±0.15 logMAR. Exclusion of this observation from the comparison of standard deviations across the three degrees of defocus did not compromise the strength of the evidence against the null hypothesis of equal standard deviations ($p=0.00005$).

Figure 12.3.1. Quantile-normal plot to assess +0.50 defocus data for departures from normality



13. DISCUSSION

The results of this study suggest that optical defocus strongly influences the precision of visual acuity measurements. This finding is consistent with the work of Carkeet[213] who demonstrated a significant flattening of the frequency-of-seeing curve (as described using Probit size[214]) for optical defocus of 1.00D or 2.00D. The data show a considerable reduction in precision even for degrees of optical defocus as small as 0.50 dioptres, a degree of blur which is consistent with an acuity of approximately 6/9 (+0.18 logMAR)[215 216]. This study has estimated precision, in terms of the 95% TRR, to be ±0.10 logMAR for normal subjects wearing a full refractive correction. This figure appears consistent with the results of Elliott et al[74] (±0.07 logMAR), Bailey et al[80] (±0.09 logMAR) and Arditi et al[79] (±0.09 logMAR), all studies featuring normal subjects wearing their full refractive correction.

Review of published estimates of acuity test precision does not suggest a strong relationship between ocular abnormality and reduced precision (see Table 11.1). At least three studies have reported on the effects of ocular abnormality on precision. Vanden Bosch et al[78] found no difference in precision between patients with various forms of maculopathy (mean VA +0.32 logMAR) and age-matched normals (mean VA -0.11 logMAR). Blackhurst et al[208] reported that, for subjects with macular degeneration, a smaller percentage of test-retest measurements fell within a given range as compared with normals. This difference did not, however, attain statistical significance. Elliott et al[74] found the 95% TRR to be wider for subjects with cataract than those without by ±0.02 logMAR. This difference was not tested statistically but equates to a difference of plus or minus one ETDRS letter. In all three studies, subjects were tested wearing their full refractive correction. It has been suggested that small differences in estimates of precision such as those demonstrated in these studies may be due to the truncated nature the of measurement scale which may artificially reduce TRV for higher (better) acuity scores[83]. The present study employed a 6.3 metre testing distance to avoid such effects. There is some

evidence that precision is similar for abnormal subjects and acuity-matched normals[217]. Thus it appears that a relationship between precision and ocular abnormality may exist, but that it is likely to be weak compared with that between precision and uncorrected refractive error.

The considerable loss of precision with even small degrees of optical defocus has important implications for both clinical practice and clinical research, because the width of the 95% TRR determines the ability of a test to detect change (see section 2.9). Based upon these data, for individuals whose vision is being monitored, the detection of a deterioration in vision will be delayed if refractive errors as small as 0.50 dioptres are left uncorrected. For a clinical research study using visual acuity as a primary outcome measure, an unnecessarily large sample size would be required to detect a given degree of change if visual acuity measurements are not conducted with a full refractive correction.

The results of this study should be interpreted in the light of evidence which suggests that the reduction in visual acuity associated with optical defocus may be less pronounced following a period of adaptation[218 219]. It is therefore conceivable that any reduction in precision associated with optical defocus may be less pronounced for subjects who are habitually in a state of defocus than for those who are not. A separate study would be required to elucidate the strength of any relationship between adaptation time and precision. In the absence of further evidence, and in the light of the apparent strength of the relationship between defocus and test precision, it would seem prudent for clinicians and clinical researchers to measure visual acuity with a full refractive correction wherever possible. Additional investigation is required to establish whether the deleterious effects of optical defocus on precision are mitigated by increased adaptation time. Further research is also warranted to determine whether the effect of optical defocus upon precision is influenced by the presence of eye disease, and if so whether or not the effect varies with different forms of ocular abnormality.

Link section

Sections 5 to 13 of this thesis have considered some of the factors which determine the precision of an visual acuity test. The greater the precision with which acuity can be measured (i.e the less the amount of test-retest variability or TRV), the earlier change may be detected in individuals, and the smaller the difference which may be detected between groups (see section 2.9). When monitoring a patient over time, a clinician is required to disregard apparent changes in acuity resulting from TRV alone, but to recognise any change which reflects a genuine alteration in clinical status. To this end, it would be desirable to have a cut-off or 'change-criterion' against which measured change could be judged. The width of the 95% TRR as determined using the methods of Bland and Altman (see section 4.1.2) has been widely advocated as a cut-off against which measured change may be judged[74 78-81 84 212 220]. Various estimates of the width of the 95% TRR for the ETDRS chart have been published ranging from ±0.07 to ±0.19 logMAR (see Table 11.1).

Employing the 95% TRR as a change-criterion ensures that, in the absence of true clinical change (and providing various assumptions are met – see section 4.1.4), only 5% of measured differences will exceed this level. In other words the width of the 95% TRR tells us what change-criterion is required to fix specificity at a level of 95%. However, another important characteristic of a diagnostic test, sensitivity, is not considered (see section 4.1.6 for definitions of sensitivity and specificity). To the author's knowledge, no study has attempted to evaluate visual acuity test performance in terms of sensitivity, or investigated how sensitivity to change and specificity are influenced by the change-criterion used. This may be due in part to the fact that measuring sensitivity to change for a visual acuity test is difficult because we lack an independent 'gold standard' method of establishing whether true clinical change has occurred and if so, precisely how much. It is possible to simulate changes in acuity e.g. through image degradation and see how sensitive a visual acuity test is to this change. However, we are again faced with the problem that we

do not know what degree of true underlying change in vision we have induced, and without this information, the estimate of sensitivity is meaningless. For the same reason, the introduction of retinal image blur through optical defocus is not a practical method of simulating a change in acuity (in addition to the fact that precision, and therefore the detection of change, varies with the degree of optical defocus – see sections 12 and 13).

To overcome this problem, the following study utilises one of the advantages of contemporary logMAR visual acuity charts, viz. the difficulty of the test can be adjusted with absolute precision by altering the viewing distance. For example, the 0.7 logMAR letters on a chart viewed at 4 metres are precisely equivalent to the 1.0 logMAR letters viewed from 8 metres. Hence, although we cannot alter an individual's acuity by an exact amount, we can alter, with great precision, the difficulty of the task which makes up the test. Another welcome advantage of logMAR visual acuity charts is that the test's scale increment (which influences its ability to detect change) is independent of viewing distance.

14. MEASURING THE SENSITIVITY TO CHANGE OF LOGMAR VISUAL ACUITY MEASUREMENTS

14.1. AIM

Primary Aims

- To assess the performance of the ETDRS logMAR chart in terms of sensitivity to change and specificity.
- To determine whether the approach of repeating and averaging offers improved test performance when measured in terms of sensitivity to change and specificity.
- To determine the effect of repeating and averaging on the width of the 95% TRR using a larger number of subjects than was employed in a previous study described in section 8.

Secondary Aims

- To determine whether the increased luminance of a TFT screen improves the accuracy of the PC-test.
- To determine whether PC-test acuities are influenced by pixelisation effects.
- To determine whether the intra-test variability within a PC-test can be used to create an individualised change-criterion and hence improve the detection of change.

14.2. METHODS

14.2.1. *Subjects*

50 subjects were recruited from the staff of Moorfields Eye Hospital NHS Trust, London, UK; and the students of The City University, London, UK. Subjects gave verbal consent following a detailed explanation of the testing procedure. Inclusion criteria were:

- Age under 50 years,
- Absence of any ocular abnormality, including media opacity,
- Ability to understand and comply with the testing protocol, and
- Snellen acuity of +0.18 logMAR (approximate Snellen equivalent 6/9) or better.

One eye of each subject was assessed (see section 4.2.5). Where both eyes met the inclusion criteria, the right eye was used as the study eye.

14.2.2. *Refractive correction*

All subjects wore their full distance refractive correction as established by formal subjective refraction (see section 4.2.4) at a testing distance of 8 metres immediately prior to the study measurements being taken.

14.2.3. *Equipment*

The ETDRS chart

The ETDRS chart is described in full in section 2.13. For this study, exact replicas of the 3 ETDRS charts (charts 1, 2, and R) were produced by a specialist printing house. These charts differed from the originals only in that they each represented a mirror image of the original. This was done such that the charts could be viewed through a mirror. The charts were mounted on clear plastic rather than the translucent plastic on which the standard ETDRS charts are printed upon. This resulted in a higher background luminance compared with the original chart (520 cd/m^2 versus 300cd/m^2), but the same level of contrast (98%).

The PC-test

For this study, an 18 inch TFT flat panel display ('Syncmaster 800 TFT', Samsung Ltd) was used to display the letter stimuli as opposed to the CRT screen used in

previous studies (see sections 8 and 10). This screen has a native resolution of 1280x1024 pixels and a non-interlaced refresh rate of 75 Hz. The PC-test is described in full in section 7.2. The test featured 5 thresholds and a 'jump' of 5 (see section 7.2.3). The PC-test as used in this study differed from that previously described in that it displayed a mirror image of the usual display such that the test could be viewed through a mirror. The TFT display produced a background luminance of 160 cd/m^2, and a contrast of 98%.

14.2.4. *Adjustment of test distance*

The use of direct viewing of acuity tests was considered inappropriate for this study because for the longer testing distances (e.g. 8 metres) the examiner would be too far from the subject to interact with them, or to point at the chart for example in the event of the of the subject loosing their place. The subject was therefore positioned in front of an optical quality mirror mounted on a movable floor stand, with the ETDRS chart on their left side and the PC-test on their right. Floor markings were used to correctly position the mirror to create each required testing distance for each test.

14.2.5. *Testing procedure*

Each subject underwent an ETDRS acuity measurement and a PC-test acuity measurement at each of 5 different distances (4.0, 4.5, 5.0, 6.3 and 8.0 metres). An additional measurement was taken using each test at the 4 metre reference distance making a total of 12 measurements in all. The additional measurement at 4 metres was required to provide a series of paired measurements from which the width of the 95% TRR could be estimated (see section 4.1.2). This paired series of measurements also allows an estimate of specificity to be made for each test. Taking the 4 metre distance as the reference distance, each increase in distance increases the difficulty of the test by a precise amount (see Table 13.2). The increments in relative difficulty were chosen to cover the range of precision estimates reported in the literature. The 12 measurements were carried out in random order, the only proviso being that the same ETDRS chart should not be used for consecutive measurements. To minimise memory effects as a source of bias when using 3 ETDRS charts for 6 ETDRS measurements, the charts were each read once forwards, and once backwards. An earlier study has

demonstrated that reading a chart backwards does not increase difficulty of the test (see section 12.3).

14.2.6. *Scoring*

The ETDRS charts were scored using the interpolated method and an end point of a full row of errors (see section 4.2.6). The PC-test algorithm uses an identical method to calculate the acuity score following each individual acuity threshold (see section 7.2.3). For the ETDRS chart, all responses were recorded on data proformas (see section 4.2.10) and the scores calculated after the testing session.

Table 13.2. Relative increase in task difficulty with increasing viewing distance

Viewing distance (m)	Relative difficulty (logMAR)
4.0	Reference
4.5	+0.05
5.0	+0.10
6.3	+0.20
8.0	+0.30

14.2.7. *Analysis*

Categorising subjects into 'change' and 'no change'

For each test, and for each degree of simulated change (or for no simulated change where both measurements were conducted at the same distance), subjects were categorized as having changed if the measured acuity change exceeded the relevant change-criterion, otherwise they were categorized as having not changed. The change-criteria investigated were those which could be identified from the published estimates of the 95% TRR (Table 11.1). Sensitivity and specificity are defined in section 4.1.6. For the purposes of this study, these definitions can be restated as follows:

Sensitivity to change

Sensitivity was defined as the percentage of individuals having undergone a simulated change whom the test correctly identified as having changed.

Specificity

Specificity was defined as the percentage of individuals who, on the basis of the two test results at 4 metres, were not identified as having changed.

14.3. RESULTS

Of the subjects recruited, 15 (30%) were emetropic, 29 (58%) myopic, and 6 (12%) hypermetropic. The refractive error (spherical equivalent) ranged from -8.00 to +2.00 with a mean of -1.51 dioptres. 8 subjects (16%) had astigmatism ≥0.75 dioptres (maximum 1.75 dioptres).

14.3.1. *ETDRS performance*

Prior to estimating the precision of the ETDRS test, the paired measurements taken at the 4 metre reference testing distance were subtracted from one another for each subject. The resultant distribution of differences was assessed using the Shapiro-Wilk W-test (see section 4.1.5) for evidence of departure from a normal distribution. There was no evidence at the 5% level for a departure from a normal distribution (W=0.98, p=0.55). The methods of Bland and Altman were therefore used to estimate the width of the 95% TRR (see section 4.1.2). The estimated 95% TRR for the ETDRS chart was ±0.11 logMAR. This figure was added to the list of previously published estimates of 95% TRR in Table 11.1.

Table 14.3.1 shows the sensitivities and specificities achieved using the ETDRS chart for the various change-criteria listed previously in Table 11.1 (plus the internal change-criterion calculated from the paired measurements taken at 4 metres in this study). As would be expected, a trade-off exists between sensitivity and specificity. For a given degree of simulated change, as the size of the change-criterion increases, sensitivity reduces whilst specificity increases. No change-criterion achieves a sensitivity and specificity in excess of 90% for a simulated acuity change of 0.10 logMAR or less. For larger simulated changes (0.20 or 0.30 logMAR), high levels of sensitivity and specificity were achievable using the internally derived 95% TRR (±0.11 logMAR).

Table 14.3.1. Specificity and sensitivity for the ETDRS charts for each degree of simulated change using change-criteria derived from published 95% TRR figures.

Change-criterion	Sensitivity (95% CI)				Specificity
	0.05 change	0.10 change	0.20 change	0.30 change	no change
0.07[74 78]	58 (43-72)	60 (45-74)	100 (93-100)*	100 (93-100)*	78 (64-88)
0.08[211]	44 (30-59)	50 (36-65)	100 (93-100)*	100 (93-100)*	78 (64-88)
0.09[79]	44 (30-59)	50 (36-65)	100 (93-100)*	100 (93-100)*	88 (76-95)
0.10[80]	18 (9-31)	38 (25-53)	100 (93-100)*	100 (93-100)*	88 (76-95)
0.11[Δ]	18 (9-31)	38 (25-53)	100 (93-100)*	100 (93-100)*	96 (86-100)
0.15[ΔΔ]	4 (0-14)	18 (9-31)	84 (71-93)	100 (93-100)*	100 (93-100)*
0.16[81]	4 (0-14)	12 (5-24)	66 (53-79)	100 (93-100)*	100 (93-100)*
0.18[ΔΔΔ]	2 (0-11)	4 (0-14)	54 (40-68)	98 (89-100)	100 (93-100)*
0.19[84]	2 (0-11)	4 (0-14)	54 (40-68)	98 (89-100)	100 (93-100)*

All decimal figures in bold are logMAR
[Δ] – indicates the estimated 95% TRR from the present study
[ΔΔ] – indicates the estimated 95% TRR from the study described in sections 10 and 11
[ΔΔΔ] – indicates the estimated 95% TRR from the study described in sections 5 and 6
* indicates the use of a one-sided 97.5% confidence interval

14.3.2. *Accuracy of the PC-test as compared with the ETDRS chart*

To assess accuracy, the ETDRS 4 metre measurements were subtracted from those for the PC-test. Linear regression suggested no acuity-related bias between measurements produced by the two tests (t=-1.20, p=0.236, see section 4.1.6). Accuracy was therefore described using the mean of the differences between the paired measurements taken using the two tests. The mean difference between the two tests was -0.010 logMAR which is equivalent to half an ETDRS letter. The 95% confidence interval for the mean included zero (95% c.i. -0.027 to 0.007 logMAR) suggesting no statistically significant bias (section 4.1.1).

14.3.3. *Validating the distance manipulation approach*

The manipulation of viewing distance to simulate a change in acuity assumes that the theoretical relationship between the viewing distance and measured acuity is a valid

one. If this assumption is valid then, on average, an acuity measured using a 5 metre viewing distance will be 0.10 logMAR worse than one measured from 4 metres, and an acuity measured using an 8 metre viewing distance will be 0.30 logMAR worse than one measured from 4 metres. We can therefore modify the method used to assess accuracy (see section 4.1.1), to test the validity of this assumption. For each of the viewing distances other than the reference standard, the measured acuities were subtracted from those measured at the reference standard 4 metre distance. This gives four series of difference measurements for the ETDRS test, and four for the PC-test (one for each degree of simulated change). Table 14.3.2 shows the mean difference for each test and each viewing distance alongside the theoretically expected difference. In every case for the ETDRS chart, the mean difference is close to the expected difference, and the 95% confidence interval for the mean includes the expected difference. This suggests that the relationship between testing distance and measured acuity is as predicted theoretically (at least for testing distances of between 4 and 8 metres).

Table 14.3.2. Difference in acuity with increasing test distance relative to standard 4 metre distance – expected versus observed.

Viewing distance (metres)	Expected difference (logMAR)	Observed mean difference (95% CI) (logMAR)	
		ETDRS	PC-test
4.5	-0.050	-0.059 (-0.079 to -0.039)	-0.050 (-0.062 to -0.038)
5.0	-0.100	-0.093 (-0.109 to -0.077)	-0.098 (-0.112 to -0.085)
6.3	-0.200	-0.196 (-0.209 to -0.182)	-0.198 (-0.212 to -0.184)
8.0	-0.300	-0.305 (-0.321 to -0.288)	-0.302 (-0.315 to -0.289)

14.3.4. *Assessing the influence of pixelisation effects on PC-test acuity scores*

Having demonstrated using the ETDRS data that the relationship between testing distance and visual acuity is as predicted theoretically, Table 14.3.2 may be used to assess whether acuities measured using the PC- test are influenced by pixelisation effects. If we consider two acuity stimuli which subtend the same visual angle at the eye, but one of which is displayed on a monitor at a viewing distance of 4 metres and

one is displayed on a monitor at 8 metres. The stimulus displayed at 8 metres will have twice the number of pixels per unit area compared with the stimulus viewed at 4 metres. Therefore, if pixelisation effects exert a detrimental effect on the legibility at 4 metres, then the PC-test will underestimate acuity at 4 metres. Accordingly, the difference between acuities measured at 4 metres and 8 metres would be less than that predicted by theory. Table 14.3.2 shows that PC-test acuities measured at 8 metres are, on average, almost exactly 0.3 logMAR worse than those measured at 4 metres. This suggests that pixelisation effects do not influence PC-test measurements on normal subjects at a testing distance of 4 metres.

14.3.5. *Comparing the performance of the ETDRS chart with that of the PC-test*

Precision in terms of the 95% Test-Retest Range (TRR)

Based upon the paired series of measurements taken using the ETDRS chart, the precision of the ETDRS chart in terms of the 95% TRR was estimated to be ±0.11 logMAR (section 12.3.1). Prior to using the methods of Bland and Altman to estimate the precision of the PC-test such that the performance of two tests can be compared, the distribution of differences for this test was assessed for evidence of non-normality (see section 4.1.5).

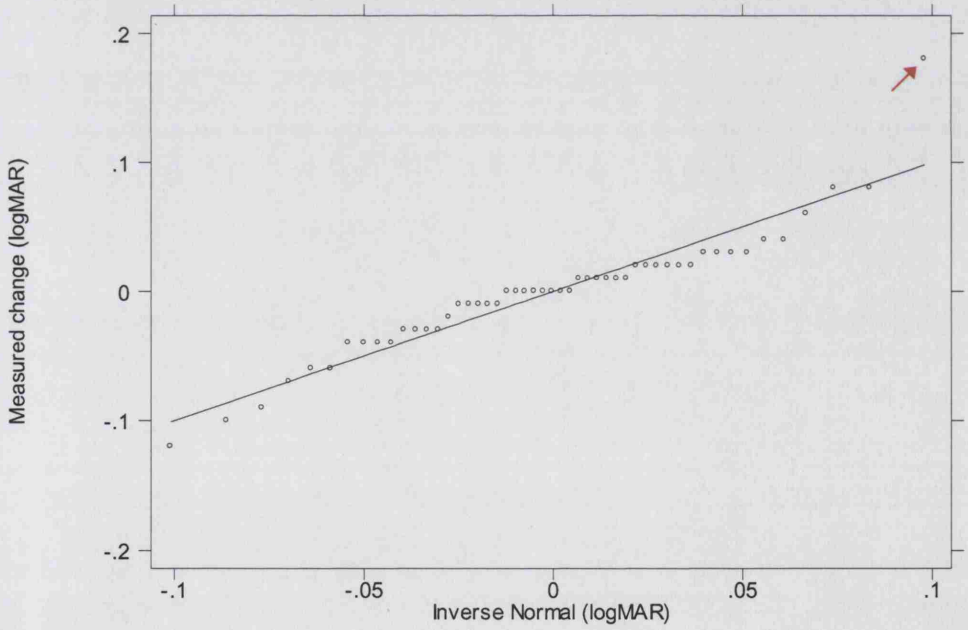
Table 14.3.3. Assessing the distribution of test-retest data for normality

Test	W*
ETDRS	0.980 (p=0.552)
PC-test	0.932 (p=0.006)

* - Shapiro-Wilk W-test

Table 14.3.3 shows the results of the Shapiro-Wilk test as applied to the distribution of differences between test and retest for the ETDRS chart (as reported in section 14.3.1) and the PC-test. In contrast to the results for the ETDRS chart, there is strong evidence of a departure from normality for the PC-test ($p<0.01$). In order to further assess this distribution, a quantile-normal plot was generated (Fig 14.3.1).

Figure 14.3.1. Quantile-normal plot for the PC-test



Most of the points in Fig 14.3.1 fall close to the oblique line indicating good conformity to a normal distribution. The exception is the point indicated by a red arrow in the upper right hand corner of the plot. This outlying point related to subject 1 for whom the difference between ETDRS test and retest was +0.18 logMAR (as opposed to the expected value of zero). Recalculation of the acuity scores for both measurements from the original data proforma ruled out any scoring error as a cause of this outlier. However, the design of this experiment offers additional scope for

experimental error as compared with previous studies due to the fact that the mirror must be repositioned between each measurement. To rule out an error in positioning of the mirror as a cause of this outlier, the acuity scores for this subject and the sequence in which they were measured was examined (see Table 14.3.4).

Table 14.3.4. PC-test measurements for subject 1

Test distance (metres)	logMAR score (PC-test)	logMAR score (ETDRS)
4.0	-0.03 (8)	-0.24 (9)
4.0	-0.21 (4)	-0.20 (10)
4.5	-0.19 (11)	-0.12 (3)
5.0	-0.11 (2)	-0.06 (6)
6.3	-0.03 (7)	-0.04 (5)
8.0	+0.05 (12)	+0.10 (1)

The numbers in parentheses refer to the order in which the measurements were performed

Inspection of Table 14.3.4 reveals two notable facts. Firstly, of the two 4 metre PC-test measurements which were subtracted to produce the outlying difference value, it is the -0.03 logMAR value which appears least consistent with both the other measurements taken using the PC-test, and the ETDRS scores. Also, the measurement taken immediately before the erroneous one was taken using the same test (the PC-test) at a distance of 6.3 metres. The scores from these two measurements are identical (measurements 7 and 8). Had the examiner neglected to reposition the mirror between measurements 7 and 8, the viewing distance would have been 2.3 metres longer than expected, and the measured difference between the two 4 metre PC-test scores, approximately 0.2 logMAR more positive than expected. As the outlying difference measurement was calculated to be +0.18 logMAR rather than zero, it appears possible that experimental error was responsible for this outlying point. This finding raised sufficient doubt as to the validity of the measured difference between test and retest for the PC-test and subject 1 to warrant excluding this subject from the forthcoming analysis.

Recalculating the Shapiro-Wilk W-statistic following removal of the data for subject 1 gave no evidence of a departure from normality (W=0.97, p=0.158).

Table 14.3.5. Precision of the ETDRS and PC-test (excluding subject 1)

Test	95% TRR	F*
ETDRS	±0.11 logMAR	0.525 (p=0.028)
PC-test	±0.08 logMAR	

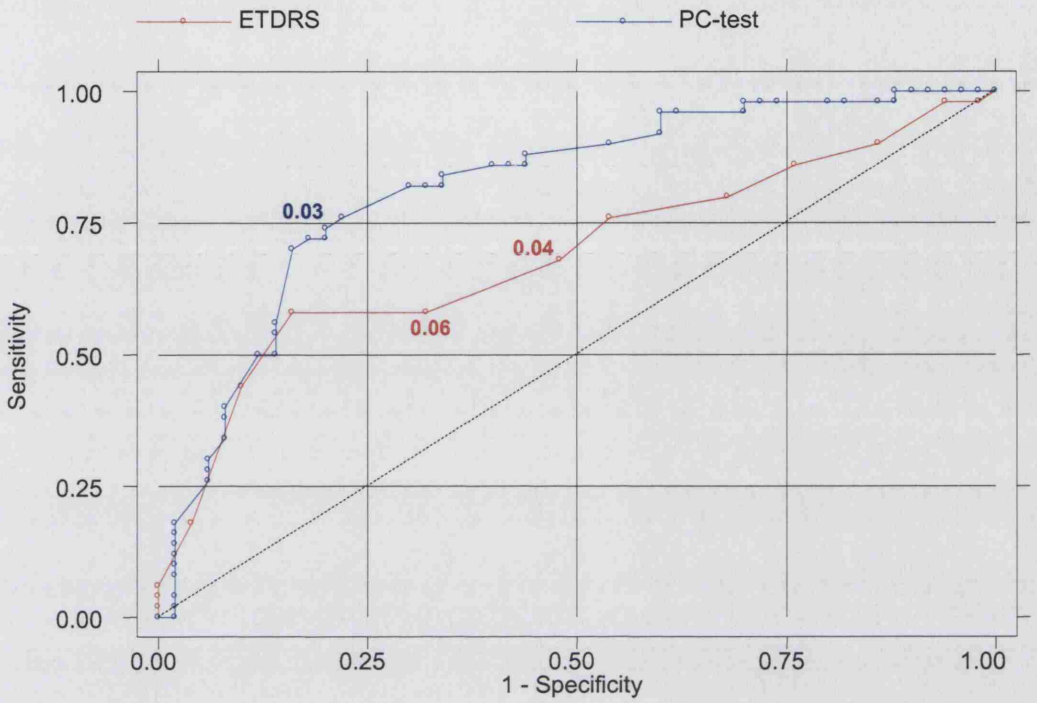
* – F-test for a difference in variance (see section 4.1.6)

In the absence of any evidence of non-normality, the methods of Bland and Altman were deemed appropriate to estimate the precision of both tests (section 4.1.2). Table 14.3.5 shows the precision of the ETDRS and PC-tests in terms of the 95% TRR following removal of subject 1 from the analysis. The width of the 95% TRR for the PC-test was ±0.03 logMAR narrower than that of the ETDRS chart. This difference equates to 1½ ETDRS letters and was significant at the 5% level.

Comparing the performance of the ETDRS and PC-tests in terms of sensitivity and specificity

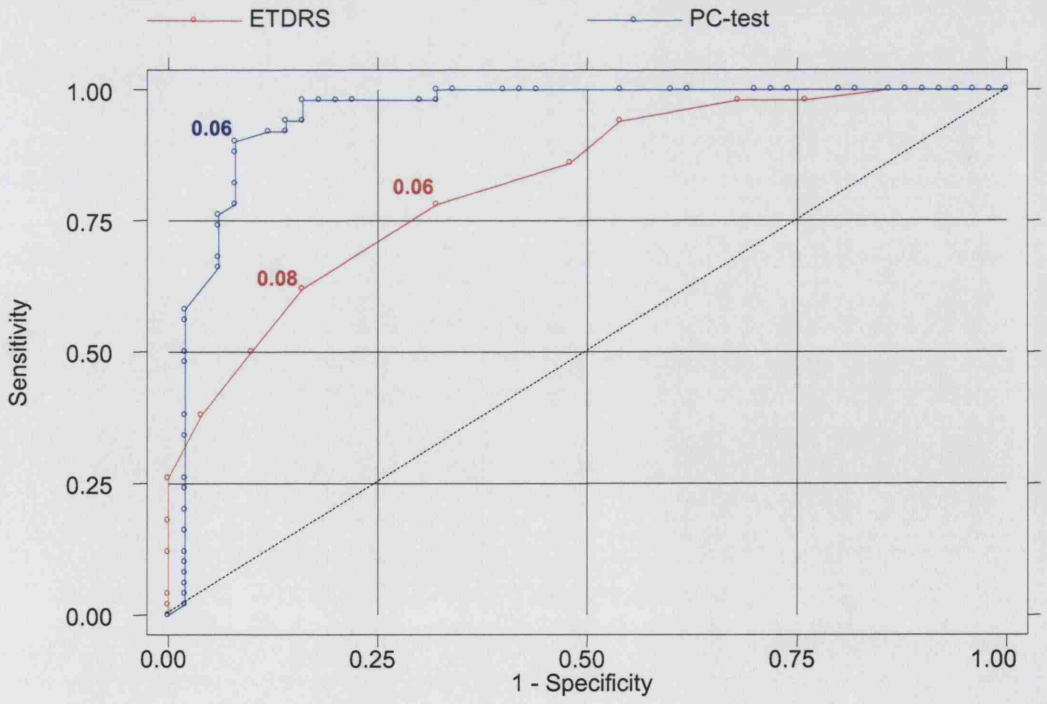
To compare the sensitivities and specificities of the ETDRS chart and the PC-test, receiver operator characteristic (ROC) curves were plotted for each degree of simulated change and each test (Figs 14.3.2-5). An ROC curve allows an overview of performance in terms of sensitivity and specificity over the whole range of possible change-criteria (see section 4.1.6).

Figure 14.3.2. ROC curves for a simulated change of 0.05 logMAR



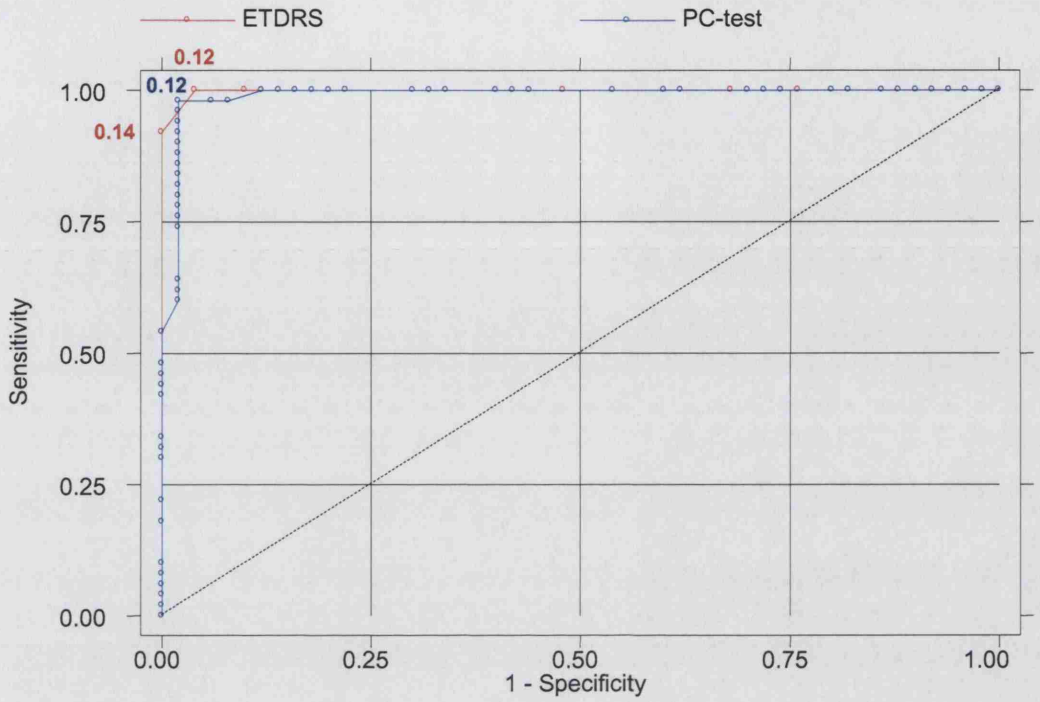
Figures in blue and red show the appropriate change-criterion

Figure 14.3.3. ROC curves for a simulated change of 0.10 logMAR



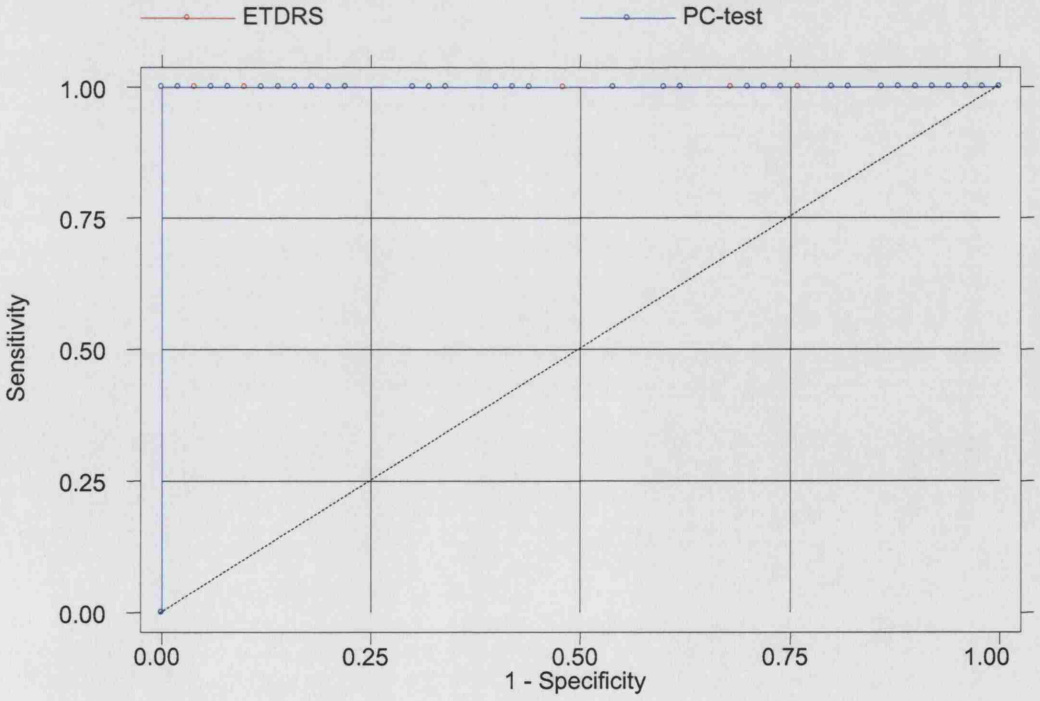
Figures in blue and red show the appropriate change-criterion

Figure 14.3.4. ROC curves for a simulated change of 0.20 logMAR



Figures in blue and red show the appropriate change-criterion

Figure 14.3.5. ROC curves for a simulated change of 0.30 logMAR



Both tests show a progressive improvement in diagnostic performance as the size of simulated change in acuity becomes progressively larger. This improvement in performance is indicated by a progressive shift of the ROC curves towards the upper left corner of the plot. For simulated changes of 0.05 logMAR (Fig 14.3.2) and 0.10 logMAR (Fig 14.3.3), the ETDRS fails to achieve a combined sensitivity and specificity of 75%, whereas the PC-test achieves combined sensitivity and specificity in excess of 75% for a simulated change of 0.05 logMAR (Fig 14.3.2), and 90% for a change of 0.10 logMAR (Fig 14.3.3). Evidence of a difference in area under the ROC curve was sought using the chi^2 test. Comparison of the areas under the ROC curves for the two tests (Table 14.3.6) suggests that the diagnostic performance of the PC-test is significantly better than that of the ETDRS chart for simulated changes of 0.05 logMAR (p=0.04) and 0.10 logMAR (p=0.006).

Table 14.3.6. Comparison of the areas under the ROC curve for the ETDRS and PC-test

Degree of change (logMAR)	Area under the ROC curve		χ^2
	ETDRS	PC-test	
0.05	0.69 (0.58 – 0.80)	0.81 (0.73 – 0.90)	4.23 (p=0.040)
0.10	0.82 (0.74 – 0.90)	0.95 (0.90 – 1.00)	7.45 (p=0.006)
0.20	1.00 (0.99 – 1.00)	0.99 (0.97 – 1.00)	0.75 (p=0.385)
0.30	1.00 (1.00 – 1.00)	1.00 (1.00 – 1.00)	0.00 (p=N/A)

$\chi 2$ – Chi^2

For a simulated change of 0.20 logMAR (Fig 14.3.4) the diagnostic performance of both tests was extremely good with combined sensitivity and specificity in excess of 95%. For a 0.30 logMAR change (Fig 14.3.5) both tests achieved an area under the test of unity indicating perfect diagnostic performance.

14.3.6. *Can a measure of intra-test variability be used to further improve the performance of the PC-test?*

The data from this study were utilised in an attempt to determine whether an estimate of an individual subject's own variability can be used to detect change more reliably than a 'global' change-criterion (such as the 95% TRR) which reflects the variability

of a group of individuals (see section 2.10.15). Because the PC-test averages 5 acuity thresholds, a measure of the spread of these 5 thresholds may be used to estimate the subject's own variability. For one of the two PC-test measurements undergone by each subject in this study, the five individual acuity scores from which the final PC-test score is derived were extracted. This was done for each subject (excluding subject 1 – see section 14.3.5) resulting in 49 sets of 5 acuity scores. For each subject, the standard deviation of the 5 acuity scores was computed.

Figure 14.3.6. Relationship between PC-test intra-test variability and test-retest difference

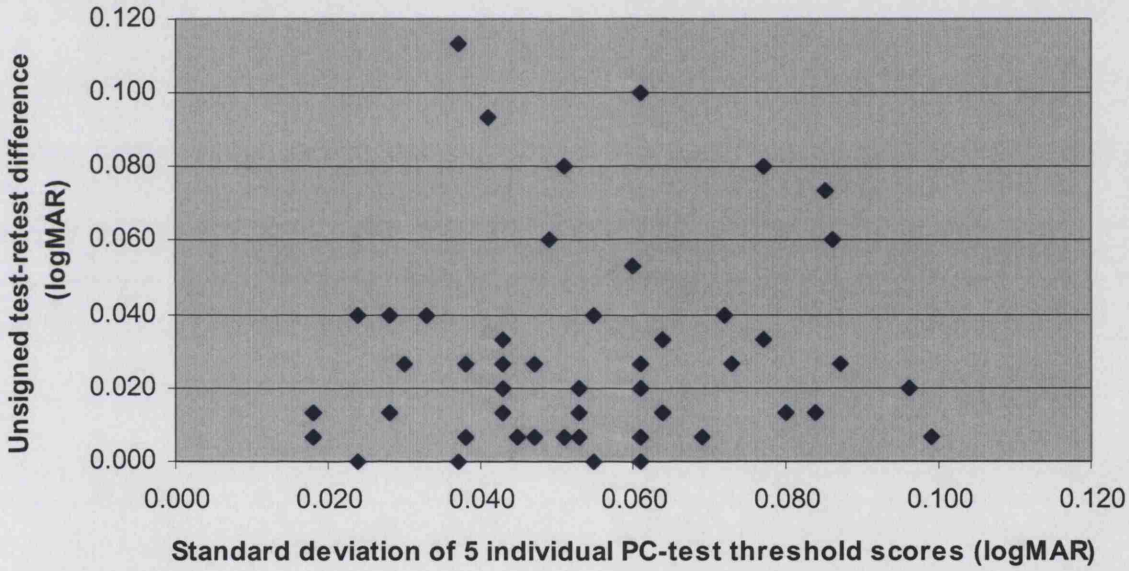


Fig 14.3.6 shows the unsigned differences between test and retest using the PC-test plotted against the standard deviation of the 5 acuity scores which make up a PC-test. The plot is not suggestive of a relationship between the two, and this was confirmed by linear regression analysis ($t=0.56$, $p=0.578$). This result suggests that a subject's variability during a PC-test measurement is not predictive of how similar the result of a second PC-test will be to the result of the first. This in turn suggests that it is not possible to use a subject's intra-test variability to detect change more reliably. In an attempt to confirm this finding, specificity and sensitivity (to simulated changes of both 0.10 and 0.20 logMAR) were recalculated using the result of a paired t-test (for a

difference in means between the 5 individual acuity scores from each PC-test measurement) as a change-criterion[A].

Categorising measured changes in PC-test acuity in to 'change' and 'no change'. If a t-test for a difference between test and retest with the PC-test yielded a p-value of less than 0.05, then the measured difference was categorised as 'change'. If the t-test yielded a p-value equal or greater than 0.05, then it was categorised as 'no change'. For comparison the measured changes were also categorised into 'change' and no 'change' using the internally derived 95% TRR as a change-criterion (±0.08 logMAR in this case).

Table 14.3.7. Performance of the PC-test using two different change criteria – the 95% TRR and the paired t-test.

Change-criterion	Specificity	Sensitivity - 0.10 logMAR change	Sensitivity - 0.20 logMAR change
t-test	90 (78-97)	65 (50-78)	98 (89-100)
95% TRR	90 (78-97)	71 (57-83)	100 (93-100)*

Numbers in parentheses indicate 95% CI, except for * which indicates one-sided 97.5% CI

Table 14.3.7 shows diagnostic performance using the two change-criteria. It can be seen that the estimates of specificity for the two methods are equal. For both degrees of simulated change the sensitivity is slightly higher when the 95% TRR is used as a change-criterion. However, the 95% confidence intervals suggest that there is no statistically significant difference between the two estimates.

[A] A paired t-test was used as this utilises the spread of the 5 acuity scores within each of 2 PC-test measurements when calculating the likelihood that the means of the two PC-test measurements are different from one another (see section 4.1.6).

15. DISCUSSION

This study has estimated the specificity and sensitivity of the ETDRS chart to varying degrees of simulated acuity change. The study has demonstrated that the change in measured acuity when viewing distance is varied between 4 and 8 metres is as predicted theoretically (which may not be the case for shorter viewing distances – see section 2.10.12). When using an internally-derived estimate of the 95% TRR (±0.11 logMAR) as a change-criterion, the ETDRS chart performed well at detecting changes of 0.2 logMAR (estimated sensitivity of 100%, 95% c.i. 93-100%; estimated specificity 96%, 95% c.i. 86-99.5%). However, for an acuity change of 0.10 logMAR, test sensitivity was poor (38%, 95% c.i. 25-53%). Furthermore, this study used normal subjects wearing their full refractive correction and all the measurements were taken by a single examiner, under identical conditions, using an interpolated scoring method. These are all factors which may have helped to maximise the degree of precision in this study, and we might therefore expect the test to perform less well in day-to-day use[A]. It should also be noted that the use of change-criteria from published studies rather than an internally derived criterion had, in some instances, a substantial impact on the estimated sensitivity and specificity of the test procedure (Table 14.3.1).

Acuities measured using the PC-test were accurate to within ½ an ETDRS letter compared with those measured using the ETDRS chart. This represents an improvement in accuracy compared with a previous study comparing the two tests in non-amblyopic subjects (see sections 8 and 9). This may have been due to the use of a TFT display which has a higher background luminance than the CRT used in the study described in section 8. Although the luminance of the reversed ETDRS chart used in this study was also greater than that of the standard ETDRS chart used in section 8, the effect of differences in test luminance is less for higher levels of

[A] The importance of interpolated scoring was demonstrated by retrospectively rescoring the ETDRS data using the line-assignment scoring method (see section 4.2.8). This increased the width of the 95% TRR to ±0.17 logMAR, and reduced sensitivity to simulated change 0.2 logMAR to 84% (95% c.i. 71-93%).

luminance (see section 2.10.9). The study also provides evidence that visual acuities measured in normal subjects are not influenced by pixelisation effects for an 18 inch screen TFT screen with 1280x1024 resolution and a 4 metre viewing distance.

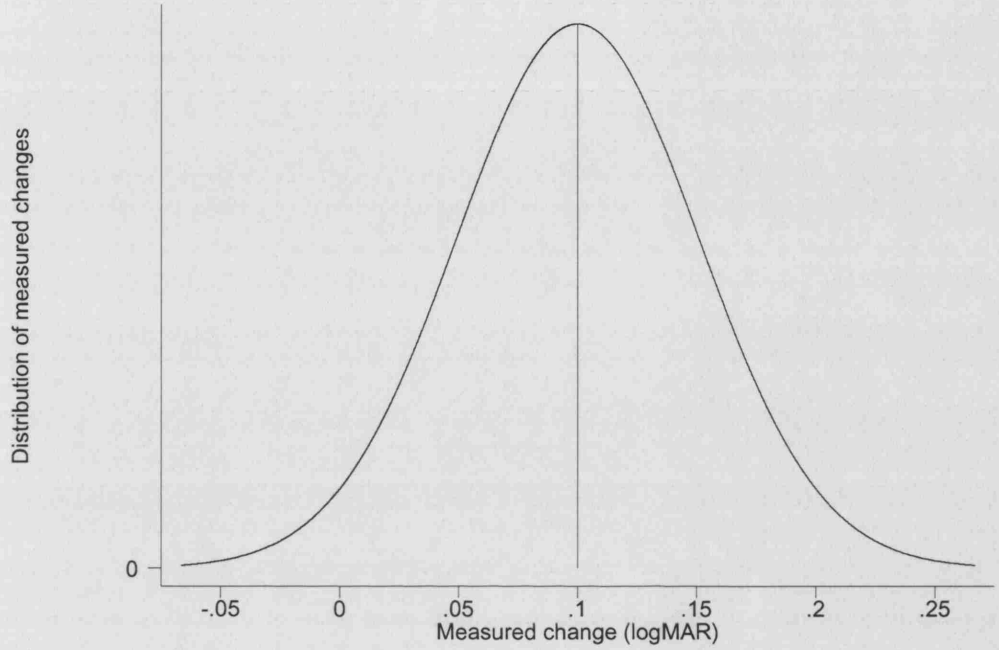
The use of repeating and averaging as incorporated in the PC-test resulted in a significant improvement in precision as compared with that of the ETDRS chart (Table 14.3.5). In turn, this provided a significant improvement in diagnostic performance for the smaller degrees of simulated change featured in this study (Figs 14.3.2 and 14.3.3, and Table 14.3.6). For simulated changes in acuity which were large relative to the estimated 95% TRR of the ETDRS chart, the diagnostic performance of the ETDRS chart was very good[A], such that the superior precision of the PC-test offered no material improvement in performance (Figs 14.3.4 and 14.3.5, and Table 14.3.6).

Our conclusions appear contrary to the those of numerous studies which have advocated the use of the 95% TRR as a cut-off against which measured changes may be judged[74 78-81 84 212 220]. The authors of these studies have correctly stated that a measured change must exceed the limits of the 95% TRR before it can be deemed 'real'. However, the failure to explicitly contrast measured change with true change creates the false impression that a test whose 95% TRR is ±0.10 logMAR is able to reliably detect true changes of ±0.10 logMAR. Due to measurement error, true changes in underlying acuity must be considerably larger than the size of the 95% TRR to ensure that the measured change exceeds the 95% TRR. This is demonstrated by considering a case where the 95% TRR is ±0.10 logMAR and hence a change-criterion of 0.10 logMAR is used. For individuals whose true underlying acuity has changed by 0.10 logMAR, the expected distribution of measured changes, assuming normality, is shown in Fig 15.1. It is clear from this figure that the change-criterion of 0.10

[A] In interpreting these results, it should be borne in mind that the predictive value of a diagnostic test depends the characteristics (sensitivity and specificity) of the test, and on the incidence of change in the population to which the test is applied[221]. If the incidence of true change is very low then even a test with a high sensitivity and specificity may result in a high proportion of test positives being false positives.

logMAR can be expected to identify 50% of these individuals as having experienced a change.

Figure 15.1. Expected distribution of measured change



Where the 95% TRR is 0.1 logMAR and the test is performed upon individuals whose true acuity has changed by 0.1 logMAR.

The results of this study have shown the PC-test to be more accurate compared with the ETDRS chart than was suggested by the study described in section 8 (see Table 8.3.1). This may be explained by the fact that the previous study utilised a CRT monitor to display the acuity stimuli, whereas the present study used a 'flat-panel' TFT screen. The difference in screen luminance may have accounted for the improved level of agreement between the two tests in this study.

Further analysis of the data also suggests that estimating each individual subject's own variability from the results of the 5 acuity thresholds within a PC-test measurement is not helpful in detecting change more reliably. This finding is contrary to the suggestions of previous authors who have advocated such an approach[143 195].

To summarise the main finding of this study; considering the performance of visual acuity tests in terms of sensitivity to change and specificity offers

an alternative perspective on the ability of these tests to detect true change. The use of the 95% TRR as a change-criterion by which to separate measured changes into true and apparent ensures a specificity of 95%[A]. However, the widespread advocacy for this approach in the literature has failed to emphasise the fact that sensitivity to true changes of equal magnitude to the 95% TRR will be approximately 50%. Sensitivity may be improved by using a change-criterion which is smaller than the minimum change sought, providing the change-criterion is still at least as large as the 95% TRR such that specificity is maintained. Improved precision will allow smaller changes in acuity to be reliably detected as a smaller change-criterion can be used without compromising specificity. This study has also demonstrated that the repetition and averaging of visual acuity measurements using the PC-test computerised system offers an improvement in precision over that obtained with the ETDRS chart. This in turn translates into a significant improvement in test performance when viewed in terms of sensitivity and specificity.

[A] Assuming that the 95% TRR used is representative of the degree of precision with which the acuity measurements are carried out.

Link section

The previous study has demonstrated that the use of the 95% TRR as a change-criterion ensures specificity of 95%, but sensitivity to true changes of equal magnitude to the change-criterion is limited to 50%. For true changes of greater magnitude than the 95% TRR, sensitivity will exceed 50% (providing the change-criterion is unchanged). The larger the change sought relative to the 95% TRR, the higher the degree of sensitivity to that change. This raises the question: how much larger than the 95% TRR must a true change be before it can be reliably distinguished from no change with e.g. sensitivity and specificity of 95%? Far from being of relevance to visual acuity alone, this issue is pertinent to the detection of change using any clinical test which is subject to measurement error. Accordingly, we sought to develop a mathematical model relating the levels of TRV and sensitivity with the magnitude of change-criterion. Such a model would allow us to calculate for a given 95% TRR (and using this as the change-criterion) the size of the smallest genuine change in acuity which can be detected with the required level of sensitivity.

16. DEVELOPING A SIMPLE MATHEMATICAL MODEL TO PREDICT THE SENSITIVITY TO CHANGE OF VISUAL ACUITY MEASUREMENTS

16.1. AIMS

- To develop a simple mathematical model to predict sensitivity to change from a knowledge of test precision (in terms of the 95% TRR).
- To compare estimates of sensitivity generated using the model with empirical data for the ETDRS chart from the previous study (described in section 14 and discussed in section 15).

16.2. METHODS

16.2.1. *Statistical model*

We assume that when visual acuity is measured using the ETDRS chart, the resultant acuity measurement reflects the individual's true acuity ± a random error. Furthermore we assume that these random errors are independent, normally distributed, random variables with mean 0 and variance σ^2 (where σ represents the standard deviation of the error in a single measurement). Under these assumptions, measured changes in status derived from the difference between 2 measurements at different points in time will be normally distributed with a mean value equal to the true underlying change in status and a variance of $2\sigma^2$. The use of the 95% TRR to derive the change-criterion results in a cut-off whose expected value is $\pm 1.96 \times \sqrt{2}\sigma$. Under the assumption of normality, we can calculate the probability that a true change of given size (d) results in a measured change greater than the change-criterion ($1.96\sqrt{2}\sigma$), and hence is detected (i.e. the sensitivity of the procedure), as:

$$P\{X > 2(1-a)\}$$

where X is a random variable, normally distributed with mean 0 and variance 1, and $a=d/(1.96\sqrt{2}\sigma)$. Note that $1.96\sqrt{2}\sigma$ represents the change-criterion so a represents the degree of change to be detected relative to the change-criterion. See Appendix B for a formal derivation of this expression.

16.2.2. *Empirical data*

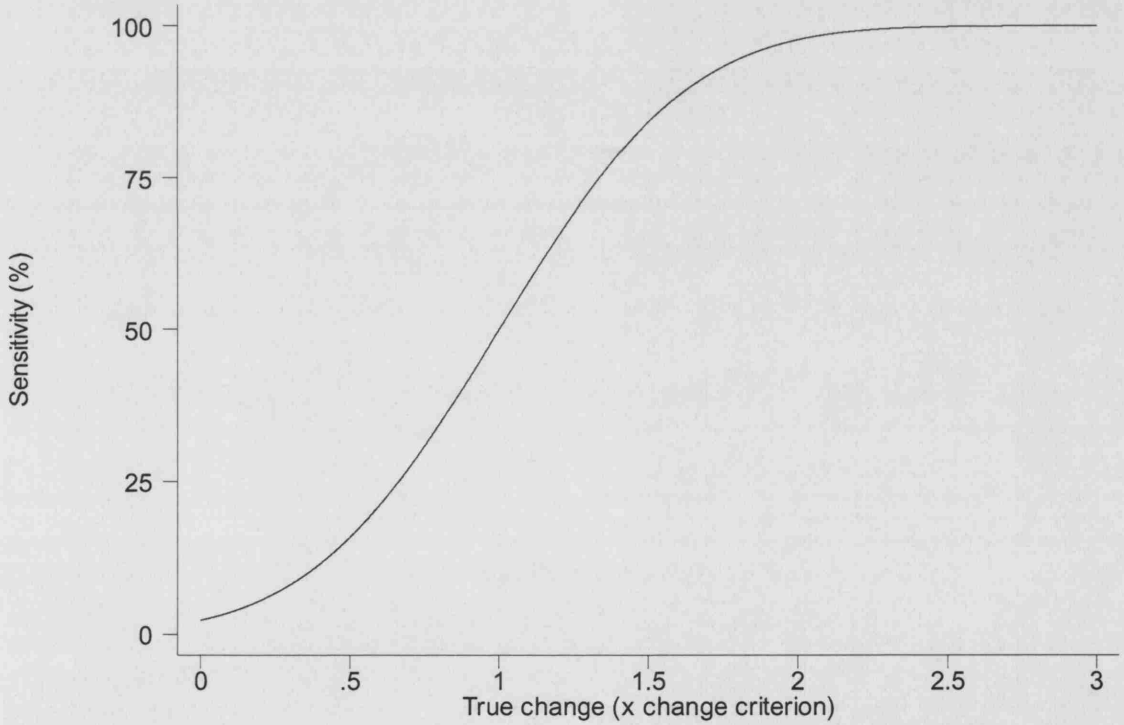
The empirical data with which the predictions using the model are compared are those collected in the previous study (see sections 14 and 15). Briefly, various degrees of acuity change were simulated in a group of 50 normal subjects by adjusting test difficulty through manipulation of the distance from which the chart is viewed. Each subject underwent acuity measurement at 5 different distances: 4.0, 4.5, 5.0, 6.3 and 8.0 metres in random order using the ETDRS logMAR chart (see section 2.13). The distance increments were chosen to simulate changes in acuity (relative to a reference measurement taken at 4 metres) of 0.05, 0.10, 0.20, and 0.30 logMAR. A second measurement was conducted at the baseline distance of 4 metres, to allow an estimate

of the 95% TRR (and σ) to be made. This internally-derived estimate of the 95% TRR was used as a change-criterion to estimate empirically the sensitivity of the procedure for identifying simulated acuity changes of 0.05, 0.10, 0.20 and 0.30 logMAR. The observed sensitivity of the procedure was compared with that predicted by the statistical model described above.

16.3. RESULTS

The predicted sensitivity of the procedure using the 95% TRR as the change-criterion is shown for varying degrees of underlying change in Fig 16.3.1. The x-axis indicates the true change in acuity as a multiple of the change-criterion of $\pm 1.96\sqrt{2}\sigma$. From Fig 16.3.1, it can be seen that the predicted sensitivity of the procedure to identify changes of similar magnitude to the change-criterion itself (i.e. 1.0 on the x-axis) is only 50%. 80% sensitivity is achieved for real changes of 1.43 times the change-criterion, increasing to 90% sensitivity for real changes of 1.65 times the change-criterion, and to 95% for real changes of 1.84 times the change-criterion.

Figure 16.3.1. Predicted sensitivity of repeated acuity measurements to detect changes in visual acuity.



From the empirical data, the distribution of test-retest differences was assessed for evidence of a departure from normality for each degree of simulated change (see Table 16.3.1). There was no evidence for such a departure at the 5% level. From the pairs of measurements made at 4.0m for each individual the standard deviation of the differences was estimated to be 0.056 logMAR ($\sqrt{2}\sigma$), and the change-criterion (based on the 95% TRR) of ±0.110 logMAR ($\pm 1.96\sqrt{2}\sigma$). Returning to Fig 16.3.1, we need, in

our setting, to multiply the numbers labelling the x-axis by 0.110 to convert them into logMAR changes. The graph then predicts a sensitivity of 50% when the true change is 0.110 logMAR (1 x the cut-off criterion), and 95.7% when the true change is 0.220 logMAR (2 x the cut-off criterion).

Table 16.3.1. Assessing the distribution of test-retest differences for normality for each degree of simulated change.

Degree of simulated change (logMAR)	W*
0.00	0.980 (p=0.552)
0.05	0.958 (p=0.074)
0.10	0.994 (p=0.995)
0.20	0.987 (p=0.866)
0.30	0.992 (p=0.979)

* - Shapiro-Wilk W-test

We compared the sensitivities predicted by the model with the empirical data observed when changes in visual acuity were simulated by manipulating the viewing distance (Table 16.3.2). In each case the observed sensitivity is close to the predicted sensitivity and includes the predicted sensitivity within its 95% confidence interval.

Table 16.3.2. Sensitivity predicted by model compared with observed sensitivity for differing degrees of real change

Simulated change (logMAR)	As multiple of change-criterion	Predicted sensitivity (%)	Observed sensitivity (%) (95% c.i.)
0.05	0.45	13.6	18 (9-31)
0.10	0.89	41.3	38 (25-53)
0.20	1.79	94.2	100 (93-100)
0.30	2.68	99.96	100 (93-100)

Table 16.3.3. Relationship between precision and predicted sensitivity for varying degrees of true change

95% TRR (logMAR)	Predicted sensitivity for varying degrees of simulated change					
	0.05 logMAR	0.10 logMAR	0.15 logMAR	0.20 logMAR	0.25 logMAR	0.30 logMAR
±0.05	50.0%	97.5%	>99.99%	>99.99%	>99.99%	>99.99%
±0.10	16.4%	50.0%	83.6%	97.5%	99.8%	>99.99%
±0.15	9.6%	25.7%	50.0%	74.3%	90.4%	97.5%
±0.20	7.1%	16.4%	31.2%	50.0%	65.2%	83.6%

17. DISUSSION

The methods of Bland and Altman have often been advocated as a way of establishing a change-criterion above which measured change is considered to reflect an important alteration in clinical status, and below which it is not[74 78-85]. Employing the 95% TRR as a change-criterion fixes test specificity at 95%. When applying a diagnostic procedure, in addition to specificity, it is desirable to know its sensitivity. Retaining the assumptions of independence and normality, a simple statistical model has been developed to predict the sensitivity of this method in detecting varying degrees of real change. The predictions from the model are compatible with empirical observations.

The quantitative findings of this study should be of interest to clinicians monitoring the visual acuity of patients, particularly if they use the 95% TRR approach to setting a change-criterion. Fig 16.3.1 and Table 16.3.3 indicate that, if our model is reasonably accurate, a clinician with a 95% TRR of ±0.15 logMAR or wider (a level of precision not uncommon among reported estimates of the TRR for the ETDRS chart – see Table 11.1) and using that cut-off, has a probability of 75% or less of detecting an individual who has experienced a 0.20 logMAR deterioration in their acuity. I.e. they will fail to identify at least one quarter of such individuals. About one in 10 patients with a deterioration of 0.25 logMAR would also be missed.

These calculations are all based on the assumption that clinicians know how precisely they measure visual acuity and hence may apply their own 95% TRR. This will often not be the case, with many clinicians using an estimate of test precision obtained from the literature. If the clinician underestimates the precision of their visual acuity test then, not surprisingly, the true specificity of the method will be lower than the assumed 95%. Perhaps less obviously, if that clinician looks up the sensitivity in Table 16.3.3 using the unrealistically narrow (incorrect) TRR they will overestimate their ability to detect real changes greater than the

change-criterion (i.e they will overestimate sensitivity). Hence an over-optimistic estimate of precision will lead to diminished specificity, allied to an over-optimistic estimate of sensitivity to change.

The model presented here makes the assumption that acuity measurements represent true underlying acuities plus or minus random errors which are independent. In reality, there may be sources of non-random errors in clinical practice, which may or may not be avoidable. A common example concerns a form of clinician-related bias, which occurs where a patient is encouraged to try harder only when their acuity measurement falls short of that expected based on the available objective evidence, but not when the clinician deems the acuity to be satisfactory. Patient-related bias may occur for example when a patient's confidence increases with each attempt at a chart, resulting in an improved acuity through an increased willingness to guess. These two sources of bias would tend to increase specificity but reduce sensitivity. It is likely that the use of strict protocols governing the testing procedure will limit these sources of non-random error. Other sources of non-random error may be harder to avoid, such as the tendency of a patient's acuity to improve as their knowledge of the letter subset used on the chart improves.

In summary, a simple statistical model has been developed to predict the sensitivity of repeated measurements of visual acuity in detecting different degrees of change. The model predictions are compatible with observed data from an empirical study (sections 14 and 15) and highlights the fact that in many settings current methods are likely to have relatively low sensitivity to detect acuity changes of clinical importance.

In the light of these findings, it would seem reasonable to suggest that in many clinical situations there is scope for improving how visual acuity measurements are conducted and interpreted. Firstly, steps should be taken to maximise precision. Ways in which this may be achieved include:

a) Utilising charts such as those based upon the Bailey-Lovie design (see section 2.13) which are based upon rigorous design principles,

b) Scoring acuities by the letter rather than by the line (see section 2.10.6),
c) Employing strict measurement protocols which feature forced-choice alternatives and strict termination rules (see section 2.10.7),
d) Avoiding uncorrected refractive errors greater than 0.25D (see sections 12 and 13).

Having taken steps to maximise precision, it is desirable for clinicians to estimate the precision of their own acuity measurements. In the light of the model presented here, this affords the clinician an awareness of the degree of sensitivity with which various degrees of real change can be detected. However, in view of the limitations of visual acuity measurements in detecting change, measured changes in acuity should be considered in the light of all available objective and subjective evidence.

While the model has been developed in the context of repeated measures of visual acuity, it is applicable to the wider range of situations in which a continuous variable is monitored for evidence of a change and in which it is reasonable to assume that measurement errors are independent and approximately normally distributed. A particular indication for the use of this model is where the width of the 95% TRR is large relative to the magnitude of the minimum clinically important change.

18. UNIFYING DISCUSSION

The results of the first of this series of studies (described in section 5) demonstrated that although relatively precise, ETDRS acuities are time consuming to measure. This may be a barrier to their uptake in routine clinical practice. An abbreviated version of the ETDRS logMAR chart allowed visual acuity to be measured in half the time required for an ETDRS measurement, and with greater precision than the Snellen chart. The study demonstrated the trade off which exists between precision and measurement time, whereby using a coarser measurement scale reduces the total measurement time (due to a reduction in the total number of letters) but at the cost of reduced test precision. The observed relationship between scale increment and precision in this study approximated that which would be expected theoretically, but the implications for measurement duration have not previously been studied empirically. In theory, the total number of letters could be reduced without compromising the scale increment by removing some of the upper rows on the chart. However, due to the fact that the main determinant of measurement duration is the number of near-threshold letters attempted (as by definition these are difficult and therefore time consuming), this is unlikely to have a substantial effect on measurement time[A].

The results of the study described in section 5 also demonstrated some of the implications of the caveats in the design of the Snellen chart. Despite having a larger total number of letters than the prototype RLM charts, Snellen acuities were less precise because the chart design is not amenable to interpolated scoring and therefore the scale increment is determined by the size interval between lines which is relatively coarse[B]. The presence of confounding variables inherent in the Snellen design also

[A] In addition, to cover the same range of acuity, such a chart would have to be used more often at a reduced testing distance which has been shown to underestimate visual acuity (see section 2.10.12).
[B] Even if interpolation were used with the Snellen chart, the need for strict scoring and termination rules would result in the measurement being as time consuming as an ETDRS measurement whilst still being less precise.

resulted in acuity-related bias whereby the chart overestimated the acuity score for subjects with good acuity, but not for subjects with poor acuity.

The author believes that an overoptimistic impression of the performance of the Snellen chart is the most notable remaining barrier to the adoption of contemporary chart designs in routine clinical practice. This study has demonstrated that when rigorous methodology is used, measured changes in Snellen acuity must reach 3 or even 4 lines in magnitude before they can be reliably attributed to a genuine change in clinical status. Paradoxically, the absence of rigorous scoring and termination rules introduces the potential for patient- and clinician-related sources of bias which may have instilled in many users of the chart an overoptimistic impression of the precision of a Snellen acuities.

The study described in section 8 describes an attempt to improve on the precision of the ETDRS chart by producing a test with a smaller scale increment. Rather than producing a chart with more than 5 letters per line, or smaller size increments between lines, a computerised method of repeating and averaging visual acuity thresholds was developed (the 'PC-test')[A]. Although the study produced two differing estimates of ETDRS precision (one for version 1 of the ETDRS chart and one for version 2[B]), PC-test acuities (which represent the average of 5 acuity thresholds with 3 letters per line and a 0.1 logMAR size interval between lines) were significantly more precise than the average precision for the ETDRS chart. Subsequent studies in amblyopic eyes (section 10) and normals (section 14) also showed PC-test acuities to be significantly more precise than ETDRS acuities. As such, the PC-test will detect a genuine change in clinical status which affects visual acuity earlier than the ETDRS chart.

[A] The use of this approach was partly to allow the test design to be varied, but also because it has been shown that subjects find a long series of near-threshold letters very tiring to attempt[190]. The author's experience is consistent with this finding.

[B] Version 'R' is identical to versions 1 and 2 except the letter difficulty is not controlled. Subsequently, the 95% TRR for version R was estimated to be ±0.10 logMAR (see section 12). This finding is not consistent with a genuine difference in precision between versions 1 and 2 which differ only in terms of the letter sequences.

Further analysis of the results of section 8 suggested that increasing the number of averaged acuity thresholds within a PC-test measurement from 1 will result in a progressive improvement in precision up to around 5 thresholds. Incorporating more than 5 thresholds is unlikely to result in a further material improvement in test precision. The effective scale increment of a PC-test featuring 5 thresholds is equivalent to that of an ETDRS chart with 15 letters per line. This provides the first empirical evidence of an upper limit to acuity test precision.

The combined results from sections 8, 10 and 14 suggest that the PC-test produces acuity scores which are accurate compared with those of the ETDRS chart. The results also suggest that a flat screen display is the preferred display medium, and that additional contour interaction may have to be incorporated if acuity deficits are not to be underestimated slightly in amblyopic eyes. Although more precise than ETDRS acuities (and therefore more able to detect change) a PC-test measurement has been shown to be considerably more time consuming than an ETDRS measurement. As such, in its current form, it is likely to be suitable only for clinical research studies in which time constraints are less than in clinical practice.

Despite being carried out under the same conditions and by the same examiner, there was a considerable degree of variation between the estimates of ETDRS precision made during this series of studies. This variation mirrors that seen in the published literature (even for studies using the same chart design and scoring rule). A review of relevant published papers suggested uncorrected refractive error as the most likely factor confounding estimates of ETDRS precision. In the absence of any existing evidence as to the relationship between optical defocus and acuity test precision, a study was designed to determine whether precision is influenced by the presence of optical defocus. The study described in section 12 showed that even degrees of defocus as small as 0.50 dioptres may have a considerable effect upon the precision of ETDRS acuities. In view of the fact that the ability of an acuity test to detect change is

strongly influenced by the precision of the test, this finding has considerable implications both for the detection of change in clinical practice, and the size of any clinical trial using visual acuity as a primary outcome measure. This finding may also explain the observation that those studies within this series in which visual acuity was measured with habitual spectacle correction produced poorer estimates of ETDRS precision than those in which any refractive error was fully corrected. It is possible the varying number of subjects with ocular abnormality may have confounded the effects of uncorrected refractive error when comparing the results of these studies. None of the studies within this series managed to demonstrate a relationship between ocular abnormality and visual acuity test precision, and there is also little evidence in the published literature to support such a relationship. However, it is likely that any study designed to seek evidence of such a relationship would have to be larger than those described herein. Accordingly, the existence of a relationship between ocular abnormality and reduced test precision cannot be ruled out but is likely to be weak compared with the relationship between optical defocus and precision.

It is apparent from the combined results of the studies featuring both the ETDRS chart and the PC-test, that the precision of the PC-test appears more consistent than that of the ETDRS chart (see table 19.1).

Table 19.1. Comparing estimates of ETDRS and PC-test precision

Study	95% TRR (logMAR)		Subjects	Refractive correction
	ETDRS	PC-test		
Section 8	±0.18*	±0.10	Mixed**	Habitual
Section 10	±0.15	±0.10	Amblyopes	Habitual
Section 14	±0.11	±0.08	Normals	Full

* - average for ETDRS charts 1 and 2

** - see Table 8.3.1

It seems likely that the precision of PC-test acuities is more independent of uncorrected refractive error and/or ocular abnormality, but unequivocal conclusions are impossible with further research. It would be highly desirable if PC-test acuities were independent of uncorrected refractive error as it would no longer be desirable to refract patients before any acuity measurement. Were PC-test acuities also independent of ocular abnormality, the size of the smallest measured change in acuity which can be attributed to a true change clinical status would become independent of the underlying level of acuity, or perhaps even the underlying condition which would simplify the monitoring of disease considerably.

Section 14 investigated the widely advocated approach of using estimates of precision in terms of the 95% TRR as a 'change-criterion' to determine whether a measured change is the consequence of a true alteration in clinical status, or just the result of test-retest variability. A novel approach was employed in which the ability of a test to distinguish real change from apparent change was considered in terms of sensitivity and specificity. Precise changes in acuity were simulated by manipulating the distance from which the chart is viewed. It was shown empirically that whilst using the 95% TRR as a change-criterion ensures a specificity of 95%, sensitivity to true changes equal in magnitude to the 95% TRR was only around 50%. This is important for visual acuity measurements because the width of the 95% TRR is relatively large compared with the size of the smallest clinically significant change. A mathematical model was created to predict, based on a knowledge of precision, the level of sensitivity with which a certain degree of true change may be detected. The model shows that a change-criterion must be at least as large as the 95% TRR to protect specificity (at a level of 95%), but sensitivity will only reach 95% for true changes which are around twice the magnitude of the 95% TRR. Consistent with considering test performance in terms of sensitivity and specificity, section 14 utilised ROC curves to demonstrate how the improved precision of the PC-test translates to a significant improvement in diagnostic performance over that of the ETDRS chart. This analysis also

demonstrated how small differences in test precision become unimportant if the size of the change sought is large relative to the size of the 95% TRR.

Implications for clinical practice and clinical research

The results of the studies described in this thesis allow the author to make various recommendations concerning the measurement of visual acuity in clinical practice and clinical research.

- The Snellen chart should be phased out in favour of charts based upon more rigorous design principles.
- An abbreviated version of the ETDRS chart may allow acuity to be measured in an acceptable period of time, and with greater precision that the Snellen chart.
- Acuity measurements should be carried out using forced-choice procedures and strict termination rules to minimise sources of bias, and allow a realistic assessment of test precision.
- All practicable steps should be taken to maximise precision. This will allow the clinician to detect change earlier, and will limit the sample size required in clinical trials which use visual acuity as a primary outcome measure.
- Steps to maximise precision include using a full refractive correction, and scoring acuities by the letter rather than by the line.
- Where possible 'internal' estimates of precision should be used to provide realistic impression of test performance.
- The precision of the ETDRS chart may be improved upon by making the scale increment finer (up to the equivalent of an ETDRS chart with approximately 15 letters per line).
- Repeating and averaging multiple thresholds of acuity is an alternative way of creating a finer effective scale increment.
- The repeating and averaging of multiple acuity thresholds can be readily performed using a personal computer. However, this process can be time consuming and may only be suitable for clinical research.
- The 95% test-retest range (TRR) may be used as a change-criterion to decide whether a measured change is due to a true change in clinical status or measurement error alone. This approach

guarantees specificity of 95% (i.e. a false positive rate of no more than 5%).

- True changes in acuity must be considerably larger than the 95% TRR before they can be reliably distinguished from no change.
- Because the magnitude of the 95% TRR for visual acuity measurements is typically large compared with the size of the smallest clinically important change, measured changes in acuity should be considered in the light of all available objective and subjective evidence.
- A simple mathematical model may be used to predict, based upon a knowledge of precision, what size of true change may be detected with a given degree of sensitivity.

Future work

- Further research is required into the effect of 'crowding bars' in abbreviated letter charts designs as well as tests which feature single letters and/or single rows. Ideally this would enable all optotype-based visual acuity tests to produce equivalent results in all forms of ocular abnormality.
- A large stratified study is required to establish unequivocally the relationship between ocular abnormality and precision. The results of this could be utilised in computerised acuity tests to produce a statistical probability that true change has occurred based upon a knowledge of the condition and the underlying acuity.
- Further work is required to assess whether the detrimental effects of optical defocus on precision are mitigated by increased adaptation time.

Future development of the 'PC-test'

Further research and development of the PC-test may improve it's usefulness as a tool for measuring visual acuity. Suggested areas for development include:

- Improved algorithms to reduce test time without compromising precision. Using a single-stimulus presentation with contour-interaction may facilitate this process. The degree of contour interaction should be such that PC-test acuities are accurate compared with ETDRS acuities for all forms of ocular abnormality.
- Incorporating statistical analysis (e.g. linear-regression analysis) such that following one or more baseline measurements, a statistical probability of true change can be computed.
- The use of statistical analysis would be aided by the development of a database which provides estimates of precision specific to a certain ocular condition and level of acuity.

BIBLIOGRAPHY

1. Westheimer G. Updating the classical approach to visual acuity. *Clin Exp Optom* 2001;84(5):258-263.

2. Donders FC. On the anomalies of accommodation and refraction. London: New Sydenham Society, 1864.

3. Bailey IL. Visual Acuity. In: Benjamin WJ, editor. *Borish's Clinical Refraction.* Philadelphia: W.B Saunders, 1998:179-202.

4. Blakemore C, Campbell FW. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J Physiol* 1969;203(1):237-60.

5. Henson DB. *Visual fields.* 2nd ed. Oxford: Butterworth Heinemann, 2000.

6. Bosworth CF, Sample PA, Gupta N, Bathija R, Weinreb RN. Motion automated perimetry identifies early glaucomatous field defects. *Arch Ophthalmol* 1998;116(9):1153-8.

7. Wall M, Ketoff KM. Random dot motion perimetry in patients with glaucoma and in normal subjects. *Am J Ophthalmol* 1995;120(5):587-96.

8. Wall M, Montgomery EB. Using motion perimetry to detect visual field defects in patients with idiopathic intracranial hypertension: a comparison with conventional automated perimetry. *Neurology* 1995;45(6):1169-75.

9. Al Khamis AR, Easterbrook M. Critical flicker fusion frequency in early chloroquine retinopathy. *Can J Ophthalmol* 1983;18(5):217-9.

10. Curran S, Wattis J. Critical flicker fusion threshold: a potentially useful measure for the early detection of Alzheimer's disease. *Hum Psychopharmacol* 2000;15(2):103-112.

11. Bennett AG, Rabbetts RB. *Clinical visual optics.* Boston: Butterworths, 1989.

12. Campbell FW, Gubisch RW. Optical quality of the human eye. *J Physiol* 1966;186:558-578.

13. Leibowitz H. The effect of pupil size on visual acuity for photometrically equated test fields at various levels of luminance. *J Opt Soc Am A* 1952;42:416-422.

14. Westheimer G. Pupil size and visual resolution. *Vision Res* 1964;4(1):39-45.

15. Brown JL. Visual acuity and form discrimination. In: Records RE, editor. *Physiology of the human eye and visual system.* Hagerstown: Haper & Row, 1979.

16. Maeda N. Wavefront technology in ophthalmology. *Curr Opin Ophthalmol* 2001;12(4):294-9.

17. Mrochen M, Kaemmerer M, Seiler T. Wavefront-guided laser in situ keratomileusis: early results in three eyes. *J Refract Surg* 2000;16(2):116-21.

18. MacRae SM, Williams DR. Wavefront guided ablation. *Am J Ophthalmol* 2001;132(6):915-9.

19. Campbell FW, Green DG. Optical and retinal factors affecting visual resolution. *J Physiol* 1965;181:576-593.

20. Applegate RA. Limits to vision: can we do better than nature? *J Refract Surg* 2000;16(5):S547-51.

21. Adler FH, Fliegelman F. Influence of fixation on the visual acuity. *Arch Ophthalmol* 1934;12(475-483).

22. Weymouth FW, Andersen EE, Averill HL. Retinal mean local sign; a new view of the relation of the retinal moasic to visual perception. *Am J Physiol* 1923;63:410-411.

23. Keesey UT. Effect of involuntary eye movements on visual acuity. *J Opt Soc Am* 1960;50(769-774).

24. Riggs LA. Visual Acuity. In: Graham CH, editor. *Vision and visual perception.* New York: Wiley, 1965.

25. Riggs LA, Ratliff F. The effects of counteracting the normal movements of the eye. *J Opt Soc Am* 1952;42:872–873.

26. Duke Elder WS. *Textbook of Ophthalmology.* St Louis: Mosby, 1938.

27. Low FN. Peripheral Visual Acuity. *Arch Ophthalmol* 1951;42:80.

28. Weymouth FW. Visual acuity - an analysis of the stimulus situation. *Am J Optom Arch Am Acad Optom* 1955;32:509.

29. Westheimer G. Visual Acuity. In: Hart WM, editor. *Adler's physiology of the eye: clinical application.* St. Louis: Mosby, 1987:531-547.

30. McDonald MA. Assessment of visual acuity in toddlers. *Surv Ophthalmol* 1986;31(3):189-210.

31. Press LJ. STYCAR ball acuity in relation to contrast and blur. *Am J Optom Physiol Opt* 1982;59(2):128-34.

32. Hecht S, Mintz EU. The visibility of single lines at various illuminations and the retinal basis of visual resolution. *J Gen Physiol* 1939;22:593.

33. Wilcox WW, Purdy DM. Visual acuity and its physiological basis. *Br J Psychol* 1933;23:233-261.

34. Hecht S, Ross S, Mueller CG. The visibility of lines and squares at high brightness. *J Opt Soc Am* 1947;37:500-507.

35. Mayer DL. Acuity of amblyopic children for small field gratings and recognition stimuli. *Invest Ophthalmol Vis Sci* 1986;27(7):1148-53.

36. Mayer DL, Fulton AB, Rodier D. Grating and recognition acuities of pediatric patients. *Ophthalmology* 1984;91(8):947-53.

37. Simons K. Visual acuity norms in young children. *Surv Ophthalmol* 1983;28(2):84-92.

38. Ryan JB. Pediatric primary care vision examination. *Optom Clin* 1996;5(2):1-34.

39. Williams D, Sekiguchi N, Brainard D. Color, contrast sensitivity, and the cone mosaic. *Proc Natl Acad Sci U S A* 1993;90(21):9770-7.

40. Westheimer G. Visual acuity. In: Hart WM, editor. *Adler's physiology of the eye.* St Louis: Mosby, 1992.

41. Harris SJ, Hansen RM, Fulton AB. Assessment of acuity in human infants using face and grating stimuli. *Invest Ophthalmol Vis Sci* 1984;25(7):782-6.

42. Harris SJ, Hansen RM, Fulton AB. Assessment of acuity of amblyopic subjects using face, grating, and recognition stimuli. *Invest Ophthalmol Vis Sci* 1986;27(7):1184-7.

43. Morgan MJ. Hyperacuity. In: Cronly-Dillon, editor. *Vision and visual dysfunction.* Boca Raton: CRC Press, 1991:87-113.

44. Enoch JM, Williams RA, Essock EA, Fendick MG. Hyperacuity: a promising means of evaluating vision through cataract. In: Chader GJ, editor. *Progress in retinal research.* Oxford: Pergamon, 1985:67-68.

45. Enoch JM, Giraldez-Fernandez MJ, Knowles R, Huang D, Hunter A, LaBree L, et al. Hyperacuity test to evaluate vision through dense cataracts; research preliminary to a clinical study. I. Studies conducted at the University of California at Berkeley before travel to India. *Optom Vis Sci* 1995;72(9):619-29.

46. Enoch JM, Knowles R. Hyperacuity test to evaluate vision through dense cataracts; research preliminary to a clinical study. II. Initial trials of the India instrument and HASP protocol at Aravind Eye Hospital, Madurai, India. *Optom Vis Sci* 1995;72(9):630-42.

47. Bennett AG. Ophthalmic test types. *Br J Physiol Opt* 1965;22:238-271.

48. Snellen H. *Letterproeven tot Bepaling der Gezichtsscherpte.* Utrecht: Weyers, 1862.

49. British Standards Institution. British Standard Test Charts for Determining Distance Visual Acuity BS 4274. London: British Standards Institution, 1968.

50. National Acadamy of Sciences - National Research Council Committee on Vision. Report of working group 39: Recommended standard procedures for the clinical measurement and specification of visual acuity. *Adv Ophthalmol* 1980;41:103-48.

51. Ferris FL, Bailey I. Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. *Ophthalmology* 1996;103(1):181-2.

52. Regan D. *Evoked potentials in psychology, sensory physiology and clinical medicine.* London: Chapman & Hall, 1972.

53. Norcia AM, Tyler CW. Spatial frequency sweep VEP, visual acuity during the first year of life. *Vision Res* 1985;25:1399-1408.

54. Bodis-Wollner I, et al. Visual association cortex and vision in man: pattern evoked occipital potentials in a blind boy. *Science* 1977;198:629.

55. Hopkisson B, Arnold P, Billingham B, McGarrigle M, Entwistle P. Visual assessment of infants: vernier targets for the Catford drum. *Br J Ophthalmol* 1991;75(5):280-3.

56. Simons K. Visual acuity and the functional definition of blindness. In: Jaeger AE, editor. *Clinical Ophthalmology*. Philadelphia: Harper & Row, 1988.

57. Cornsweet TN. *Visual perception*. New York: Academic Press, 1970.

58. Brown B, Zadnik K, Bailey IL, Colenbrander A. Effect of luminance, contrast, and eccentricity on visual acuity in senile macular degeneration. *Am J Optom Physiol Opt* 1984;61(4):265-70.

59. Elliott DB, Hurst MA. Simple clinical techniques to evaluate visual function in patients with early cataract. *Optom Vis Sci* 1990;67(11):822-5.

60. Drucker MD, Savino PJ, Sergott RC, Bosley TM, Schatz NJ, Kubilis PS. Low-contrast letter charts to detect subtle optic [corrected] neuropathies [published erratum appears in Am J Ophthalmol 1988 Jun 15;105(6):706]. *Am J Ophthalmol* 1988;105(2):141-5.

61. Koskela PU. Contrast sensitivity in amblyopia. II. Changes during pleoptic treatment. *Acta Ophthalmol (Copenh)* 1986;64(5):563-9.

62. Wolkstein M, Atkin A, Bodis-Wollner I. Contrast sensitivity in retinal disease. *Ophthalmology* 1980;87(11):1140-9.

63. Marmor MF. Contrast sensitivity versus visual acuity in retinal disease. *Br J Ophthalmol* 1986;70(7):553-9.

64. Wall M. Contrast sensitivity testing in pseudotumor cerebri. *Ophthalmology* 1986;93(1):4-7.

65. Bulens C, Meerwaldt JD, van der Wildt GJ, Keemink CJ. Contrast sensitivity in Parkinson's disease. *Neurology* 1986;36(8):1121-5.

66. Salmon JF, Carmichael TR, Welsh NH. Use of contrast sensitivity measurement in the detection of subclinical ethambutol toxic optic neuropathy. *Br J Ophthalmol* 1987;71(3):192-6.

67. Sokol S, Moskowitz A, Skarf B, Evans R, Molitch M, Senior B. Contrast sensitivity in diabetics with and without background retinopathy. *Arch Ophthalmol* 1985;103(1):51-4.

68. Regan D, Neima D. Low-contrast letter charts in early diabetic retinopathy, ocular hypertension, glaucoma, and Parkinson's disease. *Br J Ophthalmol* 1984;68(12):885-9.

69. Elliott DB. Contrast sensitivity and glare testing. In: Benjamin WJ, editor. *Borish's Clinical Refraction*. Philadelphia: W.B Saunders, 1998:203.

70. Appelboom TM. A history of vision screening. *J Sch Health* 1985;55(4):138-41.

71. World Health Organisation (WHO). Blindness information collected from various sources. Epidemiological and Vital Statistics Reports. Geneva: WHO, 1966:437.

72. Charman WN. Visual standards for driving. *Ophthalmic Physiol Opt* 1985;5(2):211-20.

73. Last JM. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press, 2001.

74. Elliott DB, Sheridan M. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt* 1988;8(4):397-401.

75. Ichikawa H, Imaizumi K, Tazawa Y, Obara Y, Ishikawa Y, Tobari I, et al. Effect of triopronin on senile cataracts. A double-blind clinical study. *Ophthalmologica* 1980;180(5):293-8.

76. Libutti G, Pannarale L. L'impiego della Bendalina nella terapie cataratta senile. Valutazione clinica dei risultati. Henuti dopo un anno di terapia. *Boll Ocul* 1985;64:645-657.

77. Bendazac lysine (Bendalina) as an anti-cataract agent: a review of basic clinical data. Proceedings of the XXV International Congress of Ophthalmology; 1987 May 4-10, 1986; Rome, Italy. Kruger, Amsterdam.

78. Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye* 1997;11:411-17.

79. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci* 1993;34(1):120-9.

80. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci* 1991;32(2):422-32.

81. Lovie Kitchin JE. Validity and reliability of visual acuity measurements. *Ophthalmic Physiol Opt* 1988;8(4):363-70.

82. Elliott DB, Sanderson K, Conkey A. The reliability of the Pelli-Robson contrast sensitivity chart. *Ophthalmic Physiol Opt* 1990;10(1):21-4.

83. Reeves BC, Wood JM, Hill AR. Reliability of high- and low-contrast letter charts. *Ophthalmic Physiol Opt* 1993;13(1):17-26.

84. Reeves B, Wood J, Hill A. Vistech VCTS 6500 charts, within- and between-session reliability. *Optom Vis Sci* 1991;68(9):728-37.

85. Reeves BC, Hill AR, Aspinall PA. The clinical significance of change. *Ophthalmic Physiol Opt* 1987;7(4):441-6.

86. Kniestedt C, Stamper RL. Visual acuity and its measurement. *Ophthalmol Clin North Am* 2003;16(2):155-70, v.

87. Jaeger E. Schriftskalen. Vienna, 1854.

88. Runge PE. Eduard Jaeger's Test-Types (Schrift-Scalen) and the historical development of vision tests. *Trans Am Ophthalmol Soc* 2000;98:375-438.

89. Küchler. Schriftnummerproben für Gesichtsleidende. Darmstadt: Küchler, 1843.

90. Snellen H. Letterproeven tot Bepaling der Gezigtsscherpte (P.W. van der Weijer 1862) cited in Bennet AG Ophthalmic test types. *Br J Physiol Opt* 1965;22:238-71.

91. Snellen H. *Test-types for the determination of the acuteness of vision.* 2nd (English) ed. Utrecht: Weyers, 1864.

92. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt* 1976;53:740-5.

93. Sloan LL. Needs for precise measures of acuity. Equipment to meet these needs. *Arch Ophthalmol* 1980;98(2):286-90.

94. Ferris FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol* 1982;94(1):91-6.

95. Green J. On a new series of letters for determining the acuteness of vision. *Trans Am Ophthal Soc* 1868;1:68.

96. Cowan A. Test letters which comply with the physiologic requirements of a visual test object. *Trans Am Ophthal Soc* 1928;26:260.

97. Ogle KN. On the problem of international nomenclature for designating visual acuity. *Am J Ophthalmol* 1953;35:909.

98. Sloan LL. New test charts for the measurement of visual acuity at near and far distances. *Am J Ophthalmol* 1959;48:807-13.

99. International classification of diseases. 9th rev (ICD-9). Geneva: World Health Organisation, 1977.

100. Blindness information collected from various sources. Epidemiological and Vital Statistics Reports. Geneva: World Health Organisation, 1966:437.

101. Sloan L. Comparison of three types of test target for the measurement of visual acuity. *Q Rev Ophthalmol* 1952;8(4):94-97.

102. Visual standards - aspects and ranges of vision loss. Sydney: International Council of Ophthalmology, 2002.

103. Dennett WS. Test type. *Trans Am Ophthal Soc* 1885;4:133.

104. Emsley HH. Visual acuity and the I.O.O. test chart. *Transactions of the Institute of Ophthalmic Opticians* 1933;II.

105. Banister H. Block capitol letters as tests of visual acuity. *Br J Ophthalmol* 1927;11:49.

106. Finkel M. An attempt to overcome the shortcomings of the conventional test chart. *Optician* 1963;146(218).

107. McMonnies CW, Ho A. Analysis of errors in letter acuity measurements. *Clin Exp Optom* 1989;79:144-51.

108. Coates WR. Visual acuity and test letters. *Transactions of the Institute of Ophthalmic Opticians* 1935;III:1935.

109. Woodruff EW. Visual acuity and the selection of test letters. *Some recent advances in ophthalmic optics*. London: Hatton Press, 1947.

110. McMonnies CW, Ho A. Letter legibility and chart equivalence. *Ophthalmic Physiol Opt* 2000;20(2):142-52.

111. McMonnies CW. Chart construction and letter legibility/readability. *Ophthalmic Physiol Opt* 1999;19(6):498-506.

112. Lebensohn JE. Snellen on visual acuity. *Am J Ophthalmol* 1962;53:152.

113. Sloan L. Measurement of visual acuity: a critical review. *Arch Ophthalmol* 1957;45:704-725.

114. Carkeet A. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. *Optom Vis Sci* 2001;78(7):529-38.

115. Johnston AW. Near visual acuity tests using Chinese characters and the logMAR principle. *Singapore Med J* 1985;26(6):448-55.

116. Khamar BM, Vyas UH, Desai TM. New standardized visual acuity charts in Hindi and Gujarati. *Indian J Ophthalmol* 1996;44(3):161-4 *LHM: This title is currently taken by UCL: Institute of Ophthalmology.

117. Al-Mufarrej MM, Abo-Hiemed FA, Oduntan AO. A new Arabic distance visual acuity chart. *Optom Vis Sci* 1996;73(1):59-61.

118. Al-Khattabi S, Oduntan AO. Arabic visual acuity chart for low vision examination. *Ophthalmic Physiol Opt* 1994;14:314-316.

119. Ruamviboonsuk P, Tiensuwan M. The Thai logarithmic visual acuity chart. *J Med Assoc Thai* 2002;85(6):673-81.

120. Wildsoet CF, Wood JM, Hassan S. Development and validation of a visual acuity chart for Australian Aborigines and Torres Strait Islanders. *Optom Vis Sci* 1998;75(11):806-12.

121. Wohlwill WF. Developmental studies of perception. *Surv Ophthalmol* 1960;28:84-92.

122. Graham FK, Berman PW, Ernhart CB. Development in pre-school children of the ability to copy forms. *Child Dev* 1960;31:339-359.

123. Rudel RG, Tember HL. Discrimination of direction of line in children. *J Comp Physiol Psychol* 1963;56:892-898.

124. Snellen H. Methods of determining the acuteness of vision. In: Oliver CA, editor. *System of diseases of the eye*. Philadelphia, 1897.

125. Sloan L. Measurement of visual acuity: a critical review. *Arch Ophthalmol* 1951;45:704-725.

126. Weber EH. De Tactu: Annotationices et Physilogicae, 1834.

127. Snell AC, Sterling S. The percentage evaluation of macular vision. *Arch Ophthalmol* 1926;54:443-461.

128. American Medical Association. Council on Industrial Health. Special Report. Estimation of loss of visual efficiency. *Arch Ophthalmol* 1955;54:462-468.

129. American Medical Association. Evaluation of permanent visual impairment. In Guides to the Evaluation of Permanent Visual Impairment, AMA, Reprinted in Physicians' Desk Reference for Ophthalmology, 1966, Medical Economics. 1984.

130. Westheimer G. Scaling of visual acuity measurements. *Arch Ophthalmol* 1979;97:327-330.

131. American National Standards Institute. International Standards Organisation, ISO/DIS 8956. Optics and optical Instruments-visual acuity testing-correlation of optotypes. 1986.

132. Velasco e Cruz AA. Historical roots of 20/20 as a (wrong) standard value of normal visual acuity. *Optom Vis Sci* 1990;67(8):661.

133. Colenbrander A. Measuring vision and vision loss. In: Tasman W, editor. *Duane's Clinical Ophthalmology*. Philadelphia: Harper & Row, 2002.

134. Helmholtz Hv, Nagel W. Minimum dimensions of perceptable objects. *Helmholtz's Treatise on Physiological Optics*. New York: Dover, 1962:32-6.

135. Wild JM, Hussey MK. Some statistical concepts in the analysis of vision and visual acuity. *Ophthalmic Physiol Opt* 1985;5(1):63-71.

136. Measurement of visual acuity - towards standardization. In vision science symposium, a tribute to Gordon G. Heath; 1988; Bloomington. Indiana University.

137. Colenbrander A. The functional vision score: a coordinated scoring system for visual impairments, disabilities and handicaps. In: Kooijman A, editor. *Low vision: Research and new developments in rehabilitation. Studies in health technology and informatics*. Amsterdam: IOS Press, 1994:552.

138. Kitchin JE, Bailey IL. Task complexity and visual acuity in senile macular degeneration. *Aust J Optom* 1981;64:235-42.

139. Raasch TW, Bailey IL, Bullimore MA. Repeatability of visual acuity measurement. *Optom Vis Sci* 1998;75(5):342-8.

140. Tinning S, Bentzon MW. A new method for exact measurements of visual acuity. Determination of threshold curves for the resolving power of the eye by computerized curve fitting. *Acta Ophthalmol Copenh* 1986;64(2):180-6.

141. Hazel CA, Elliott DB. The dependency of logMAR visual acuity measurements on chart design and scoring rule. *Optom Vis Sci* 2002;79(12):788-92.

142. Elliott D, Yang K, Whitaker D. Visual acuity changes throughout adulthood in normal healthy eyes: seeing beyond 6/6. *Optom Vis Sci* 1995;72:186-91.

143. Brown B, Lovie Kitchin J. Repeated visual acuity measurement: establishing the patient's own criterion for change. *Optom Vis Sci* 1993;70(1):45-53.

144. Lovie Kitchin JE, Brown B. Repeatability and intercorrelations of standard vision tests as a function of age. *Optom Vis Sci* 2000;77(8):412-20.

145. Haegerstrom-Portnoy G, Schneck ME, Brabyn JA. Seeing into old age: vision function beyond acuity. *Optom Vis Sci* 1999;77:141-58.

146. Nichols JJ, Mitchell GL, Zadnik K. Daily disposable vs. disposable extended wear: a contact lens clinical trial. *Optom Vis Sci* 2000;77:637-47.

147. Flom MC, Weymouth FW, Kahneman D. Visual resolution and contour interaction. *J Opt Soc Am* 1963;53(9):1026-32.

148. Fern KD, Manny RE, Davis JR, Gibson RR. Contour interaction in the preschool child. *Am J Optom Physiol Opt* 1986;63(5):313-8.

149. Leat SJ, Li W, Epp K. Crowding in central and eccentric vision: the effects of contour interaction and attention. *Invest Ophthalmol Vis Sci* 1999;40(2):504-12.

150. Rodier DW, Mayer DL, Fulton AB. Assessment of young amblyopes. Array vs. single picture acuities. *Ophthalmology* 1985;92(9):1197-202.

151. Mayer DL, Gross RD. Modified Allen pictures to assess amblyopia in young children. *Ophthalmology* 1990;97(6):827-32.

152. Simmers AJ, Gray LS, McGraw PV, Winn B. Contour interaction for high and low contrast optotypes in normal and amblyopic observers. *Ophthalmic Physiol Opt* 1999;19(3):253-60.

153. Westheimer G, Hauske G. Temporal and spatial interference with visual acuity. *Vision Res* 1975;15:1137.

154. Strong G, Woo GC. A distance visual acuity chart incorporating some new design features. *Arch Ophthalmol* 1985;103(1):44-6.

155. Raasch TW, Bailey IL. Choice of optotype and spacing affect visual acuity scores. *Invest Ophthalmol Vis Sci* 1984;25(Suppl):145.

156. Stager DR, Everett ME, Birch EE. Comparison of linear and optotype acuity in amblyopia. *Am Orthopt J* 1990;40:51-56.

157. Alexander KA. Psychophysical testing. In: Heckenlively JR, editor. *Principle and practice of clinical electrophysiology of vision.* St. Louis: Mosby, 1991:445-7.

158. Sheedy JE, Bailey IL, Raasch TW. Visual acuity and chart luminance. *Am J Optom Physiol Opt* 1984;61(9):595-600.

159. Shlaer S. The relation between visual acuity and illumination. *J Gen Physiol* 1937;21:165.

160. Ricci F, Cedrone C, Cerulli L. Standardized measurement of visual acuity. *Ophthalmic Epidemiol* 1998;5(1):41-53.

161. Musylev FI. Die Abhängigkeit dre Sehschärfe von der Entfernung des Objektes. *Acta Ophthalmol* 1937;15:216.

162. Dimmick F, Rudolph L. Checkerboard visual acuity targets: an experimental validation, Bureau of medial research, project NM-003-008 (X-423): United states naval medical research laboratory, United States naval submarine base, New London, Conn, 1948.

163. Freeman E. Accommodation, pupillary width and stimulus distance. *J Opt Soc Am* 1932;22(Dec):729.

164. Freeman E. Intensity, area, and distance of visual stimulus. *J Opt Soc Am* 1932;22(July):402.

165. Freeman E. Anomalies of visual acuity in relation to stimulus-distance. *J Opt Soc Am* 1932;22(May):285.

166. Beebe-Centre JB, Mead LC, Wagoner KS, Hoffman AC. Visual acuity and distance of observation. *J Exp Psychol* 1945;35:473.

167. Dong LM, Hawkins BS, Marsh MJ. Consistency between visual acuity scores obtained at different test distances: theory vs observations in multiple studies. *Arch Ophthalmol* 2002;120(11):1523-33.

168. Bailey IL, Lovie JE. The design and use of a new near-vision chart. *Am J Optom Physiol Opt* 1980;57(6):378-87.

169. Wolffsohn JS, Cochrane AL. The practical near acuity chart (PNAC) and prediction of visual ability at near. *Ophthalmic Physiol Opt* 2000;20(2):90-7.

170. Sloan LL, Brown DJ. Reading cards for selection of optical aids for the partially sighted. *Am J Ophthalmol* 1963;55:1187-1199.

171. Raasch TW, Rubin GS. Reading with low vision. *J Am Optom Assoc* 1993;64(1):15-8.

172. Rydberg A, Ericson B, Lennerstrand G, Jacobson L, Lindstedt E. Assessment of visual acuity in children aged 1 1/2-6 years, with normal and subnormal vision. *Strabismus* 1999;7(1):1-24.

173. Fern KD, Manny RE. Visual acuity of the preschool child: a review. *Am J Optom Physiol Opt* 1986;63(5):319-45.

174. Chandna A, Karki C, Davis J, Doran RM. Preferential looking in the mentally handicapped. *Eye* 1989;3(Pt 6):833-9.

175. Hertz BG, Rosenberg J, Sjo O, Warburg M. Acuity card testing of patients with cerebral visual impairment. *Dev Med Child Neurol* 1988;30(5):632-7.

176. Marx MS, Werner P, Cohen Mansfield J, Hartmann EE. Visual acuity estimates in noncommunicative elderly persons. *Invest Ophthalmol Vis Sci* 1990;31(3):593-6.

177. Bailey IL. Optometric care for the multi-handicapped child. *Pract Optom* 1994;5(4):158-166.

178. Dobson V, McDonald MA, Kohl P, Stern N, Samek M, Preston K. Visual acuity screening of infants and young children with the acuity card procedure. *J Am Optom Assoc* 1986;57(4):284-9.

179. Adoh TO, Woodhouse JM. The Cardiff acuity test used for measuring visual acuity development in toddlers. *Vision Res* 1994;34(4):555-60.

180. Sheridan MD, Gardiner PA. Sheridan-Gardiner test for visual acuity. *Br Med J* 1970;2(701):108-9.

181. Hyvarinen L, Nasanen R, Laurinen P. New visual acuity test for preschool children. *Acta Ophthalmol* 1980;58:507-511.

182. McGraw PV, Winn B. Glasgow Acuity Cards: a new test for the measurement of letter acuity in children. *Ophthalmic Physiol Opt* 1993;13(4):400-4.

183. Simmers AJ, Gray LS, Spowart K. Screening for amblyopia: a comparison of paediatric letter tests. *Br J Ophthalmol* 1997;81(6):465-9.

184. McGraw PV, Winn B, Gray LS, Elliott DB. Improving the reliability of visual acuity measures in young children. *Ophthalmic Physiol Opt* 2000;20(3):173-84.

185. Hered RW, Murphy S, Clancy M. Comparison of the HOTV and Lea Symbols charts for preschool vision screening. *J Pediatr Ophthalmol Strabismus* 1997;34(1).

186. Friendly DS, Weiss IP. An automated visual acuity testing computer program using the Apple II system. *Am J Ophthalmol* 1985;99(2):188-92.

187. Crossman ER, Goodeve PJ, Marg E. A computer-based automatic method for determining visual acuity. *American Journal Of Optometry And Archives Of American Academy Of Optometry* 1970;47(5):344-355.

188. Timberlake GT, Mainster MA, Schepens CL. Automated clinical visual acuity testing. *American Journal Of Ophthalmology* 1980;90(3):369-373.

189. Wong D, Plumb A. Computer automated visual acuity testing for visual screening. *Transactions Of The Ophthalmological Societies Of The United Kingdom* 1986;105 (Pt 4):498-503.

190. Bach M. The Freiburg Visual Acuity Test: automatic measurement of visual acuity. *Optom Vis Sci* 1996;73:49-53.

191. Ruamviboonsuk P, Tiensuwan M, Kunawut C, Masayaanon P. Repeatability of an automated Landolt C test, compared with the early treatment of diabetic retinopathy study (ETDRS) chart testing. *Am J Ophthalmol* 2003;136(4):662-9.

192. Moke PS, Turpin AH, Beck RW, Holmes JM, Repka MX, Birch EE, et al. Computerized method of visual acuity testing: adaptation of the amblyopia treatment study visual acuity testing protocol. *Am J Ophthalmol* 2001;132(6):903-9.

193. Beck RW, Moke PS, Turpin AH, Ferris FL, 3rd, SanGiovanni JP, Johnson CA, et al. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol* 2003;135(2):194-205.

194. McMonnies CW, Ho A. Analaysis of errors in letter acuity measurements. *Clin Exp Optom* 1996;79(144-51).

195. Woods R, Lovie-Kitchin J. The reliability of visual performance measures in low vision. *Vision Science and Its Applications, 1995 Technical Digest Series, Vol 1.* Washington, DC: Optical Society of America, 1995:246-249.

196. Camparini M, Cassinari P, Ferrigno L, Macaluso C. ETDRS-fast: implementing psychophysical adaptive methods to standardized visual acuity measurement with ETDRS charts. *Invest Ophthalmol Vis Sci* 2001;42(6):1226-31.

197. Adams AJ, Haegerstrom Portnoy G, Brown B, Jampolsky A. Predicting visual resolution from detection thresholds. *Am J Optom Physiol Opt* 1984;61(6):371-6.

198. Friedman DS, Munoz B, Massof RW, Bandeen-Roche K, West SK. Grating visual acuity using the preferential-looking method in elderly nursing home residents. *Invest Ophthalmol Vis Sci* 2002;43(8):2572-8.

199. Gordon MO, Schechtman KB, Davis LJ, McMahon TT, Schornack J, Zadnik K. Visual acuity repeatability in keratoconus: impact on sample size. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. *Optom Vis Sci* 1998;75(4):249-57.

200. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983;32:307-17.

201. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8(2):135-60.

202. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

203. Altman DG. Personal communication, 2003.

204. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965(52):581-611.

205. Altman DG. *Practical statistics for medical research.* First ed. London: Chapman & Hall, 1991.

206. Humphriss D, Woodruff EW. Refraction by immediate contrast. *Br J Physiol Opt* 1962;19:15-20.

207. Ray WA, O'Day DM. Statistical analysis of multi-eye data in ophthalmic research. *Invest Ophthalmol Vis Sci* 1985;26(8):1186-8.

208. Blackhurst DW, Maguire MG. Reproducibility of refraction and visual acuity measurement under a standard protocol. The Macular Photocoagulation Study Group. *Retina* 1989;9(3):163-9.

209. McMonnies CW. Chart memory and visual acuity measurement. *Clin Exp Optom* 2001;84(1):26-33.

210. Lenne RC, Vingrys AJ, Smith G. Using computers to test visual acuity. *J Am Optom Assoc* 1995;66(12):766-74.

211. van den Brom HJ, Kooijman AC, Blanksma LJ, van Rij G. Measurement of visual acuity with two different charts; a comparison of results and repeatability in patients with cataract. *Doc Ophthalmol* 1995;90(1):61-6.

212. Siderov J, Tiu AL. Variability of measurements of visual acuity in a large eye clinic. *Acta Ophthalmol Scand* 1999;77(6):673-6.

213. Carkeet A, Lee L, Kerr JR, Keung MM. The slope of the psychometric function for Bailey-Lovie letter charts: defocus effects and implications for modeling letter-by-letter scores. *Optom Vis Sci* 2001;78(2):113-21.

214. Finney DJ. *Probit Analysis.* 3rd ed. Cambridge: Cambridge University Press, 1971.

215. Kempf GA, Collins SD, Jarman BL. Refractive errors in the eyes of children as determined by retnoscopic examination with a cycloplegic: United States Public Health Service, 1928.

216. Crawford JS, Shagrass C, Pashby TJ. Relationship between visual acuity and refractive error in myopia. *Am J Ophthalmol* 1945;28:1220-1225.

217. Reeves BC, Harper RA, Hill AR, Wood JM. Is performance variability useful clinical information? *Doc Ophthalmol* 1991;77:128-129.

218. Mon-Williams M, Tresilian JR, Strang NC, Kochhar P, Wann JP. Improving vision: neural compensation for optical defocus. *Proc R Soc Lond B Biol Sci* 1998;265(1390):71-7.

219. Webster MA, Georgeson MA, Webster SM, Mon-Williams M, Tresilian JR, Strang NC, et al. Neural adjustments to image blur. *Nat Neurosci* 2002;5(9):839-40.

220. Manny RE, Hussein M, Gwiazda J, Marsh-Tootle W. Repeatability of ETDRS visual acuity in children. *Invest Ophthalmol Vis Sci* 2003;44(8):3294-300.

221. Altman DG. Diagnostic Tests. *Practical Statistics for Medical Research.* London: Chapman & Hall, 1991:409-19.

222. Bourne RR, Sukudom P, Foster PJ, Tantisevi V, Jitapunkul S, Lee PS, et al. Prevalence of glaucoma in Thailand: a population based survey in Rom Klao District, Bangkok. *Br J Ophthalmol* 2003;87(9):1069-74.

223. Dineen BP, Bourne RR, Ali SM, Huq DM, Johnson GJ. Prevalence and causes of blindness and visual impairment in Bangladeshi adults: results of the National Blindness and Low Vision Survey of Bangladesh. *Br J Ophthalmol* 2003;87(7):820-8.

224. Nolan WP, Baasanhu J, Undraa A, Uranchimeg D, Ganzorig S, Johnson GJ. Screening for primary angle closure in Mongolia: a randomised controlled trial to determine whether screening and prophylactic treatment will reduce the incidence of primary angle closure glaucoma in an east Asian population. *Br J Ophthalmol* 2003;87(3):271-4.

Research Support & Commercial Interests

This study was supported by the Trustees of Moorfields Eye Hospital NHS Trust (Grant No MURI 1011). The author has received no financial support from any commercial organisation. The author has no commercial interest in any of the tests featured in this thesis.

Appendix A

Summary of prospective research studies utilising the 'RLM-E chart'

The RLM-E chart is designed specifically for use in population-based epidemiological research studies. The design is based upon that of the RLM-A chart (see section 5).

The prevalence of glaucoma in Thailand

The RLM-E chart was used to measure visual acuity within a population-based survey in the Rom Klao district of Bangkok, Thailand[222]. 790 subjects aged 50 years and over underwent the following investigations: visual acuity, visual field testing, slit lamp examination, applanation tonometry, gonioscopy, and an optic disc examination after mydriasis. Acuity measurements were conducted at a testing distance of 4 metres using a 4-alternative forced choice testing paradigm, and an end point of a full row of errors. Any subject misnaming any letters on the top row of the chart was moved to 1 metre from the chart and asked to attempt the chart again. The examiners recorded the number of letters correctly named, and the distance at which the test had been attempted. The visual acuity was later calculated from this data using the interpolated scoring method (see section 4.2.9). Table A1 shows sample raw data from the survey.

Table A1. Sample raw data from the Rom Klao glaucoma survey

	C	D	E	F	G	H	I	J	K	L	M	N
1	Exam Date	Height (m)	Weight (kg)	Birthdate	Age (years)	Sex	Geriatric No.	Meds (Y/N)	E's at 4m (R)	E's at 4m (L)	E's at 1m (R)	E's at 1m (L)
2	08/12/2003	1.58	39	11/02/1937	66	M	K673	0	8	11		
3	08/12/2003	1.55	39	13/09/1941	62	F	K674	1	23	25		
4	09/12/2003	1.63	94	12/02/1938	65	M	D274	0	21	20		
5	14/12/2003	1.47	67	9	67	F	K653	1	0	19	14	
6	14/12/2003	1.62	66	05/07/2031	72	F	K733	0	18	25		
7	14/12/2003	1.57	53	05/05/1942	62	F	K721	0	21	26		
8	14/12/2003	1.7	62	02/11/1934	71	M	A067	1	5	0		10
9	14/12/2003	1.52	70	13/07/1946	58	F	K654	0	20	15		
10	14/12/2003	1.59	54.5	22/02/1950	53	F	K656	0	30	30		
11	15/12/2003	1.5	65	08/04/1937	68	F	B111	0	8	0		18
12	15/12/2003	1.41	41.5	04/01/1934	65	F	K727	0	30	26		
13	15/12/2003	1.63	51		68	M	N770	0	16	8		
14	15/12/2003	1.58	55.5	Dec-37	66	F	N771	0	15	15		
15	15/12/2003	1.55	49	23/06/1949	50	F	P823	0	17	17		
16	15/12/2003	1.44	65	31/12/1935	68	F	B112	0	10	21		
17	15/12/2003	1.51	62	1933	66	F	Q884	0	22	18		
18	15/12/2003	1.49	76	Jan-27	76	F	D244	0	0	9	17	
19	15/12/2003	1.46	56	Jun-41	62	F	F439	0	27	20		
20	15/12/2003	1.44	55	1940	59	F	B126	0	24	17		
21	15/12/2003	1.63	52.5	05/01/1946	57	F	N801	0	24	30		
22	15/12/2003	1.61	54	Jan-46	57	F	N802	0	31	22		
23	15/12/2003	1.53	69	09/08/1947	57	F	K734	0	27	24		
24	15/12/2003	1.53	69	29/10/1935	68	F	F445	1 (Timolol	0	0	0	18
25	15/12/2003	1.56	60	26/11/1934	69	M	F444	1	12	22		
26	15/12/2003	1.55	56	08/05/1942	62	F	N763	0	10	6		
27	15/12/2003	1.46	43.5	1920	79	F	D270	1 (Timoptc	25	24		
28	15/12/2003	1.71	67	1931	68	M	P857	0	15	8		
29	16/12/2003	1.54	43	18/02/1938	65	F	F436	0	27	27		
30	16/12/2003	1.54	49	1936	63	F	A049	0	14	7		

Columns K and L show the number of letters correctly named using a 4 metre testing distance. Columns M and N show the numbers of letters correctly named at 1 metre for those unable to correctly name any letters from 4 metres.

Prevalence and causes of blindness and visual impairment in Bangladeshi adults

The RLM-E chart was used to measure visual acuity within the National Blindness and Low Vision Survey of Bangladesh[223]. 12,782 adults aged 30 years and older underwent interview, visual acuity measurement, autorefraction, and optic disc examination. Visual acuity was measured using the same method employed in the Thai study by local nurses who had undergone specific training for this purpose.

Table A2. Sample raw data from the Bangladesh visual impairment survey

	A	B	C	D	E	F	G	H	I	J	K
1	SUBJECT	SEX	AGE	GLASSES	RE4M	LE4M	RE1M	LE1M			
2	1	2	45	1	27	27		.			
3	2	2	42	1	30	30	.	.			
4	3	1	30	1	30	30	.	.			
5	4	1	40	1	25	28	.	.			
6	5	2	34	1	29	29	.	.			
7	6	2	37	1	29	29	.	.			
8	7	1	43	1	30	30	.	.			
9	8	2	45	1	28	30	.	.			
10	9	1	32	1	30	30	.	.			
11	10	2	57	1	30	28	.	.			
12	11	1	60	1	28	28	.	.			
13	12	1	35	1	30	28	.	.			
14	13	2	33	1	30	30	.	.			
15	14	1	40	1	27	26	.	.			
16	15	2	60	1	29	27	.	.			
17	16	2	40	1	30	30	.	.			
18	17	2	55	1	28	30	.	.			
19	18	1	42	1	30	30	.	.			
20	19	2	35	1	30	30	.	.			
21	20	2	40	1	30	27	.	.			
22	21	2	62	1	28	30	.	.			
23	22	1	63	1	29	29	.	.			
24	23	1	34	1	30	30	.	.			
25	24	1	75	1	26	25	.	.			
26	25	2	36	1	30	30	.	.			
27	26	2	51	1	27	25	.	.			
28	27	1	47	1	25	30	.	.			
29	28	1	32	1	30	30	.	.			
30	29	2	34	1	30	30	.	.			

1608 subjects were found to have vision less than Snellen equivalent 6/12 in their better eye.

Primary angle closure glaucoma (PACG) in Mongolia

The RLM-E chart was used to measure visual acuity within a randomised controlled trial to determine whether screening and prophylactic laser iridotomy will reduce the incidence of primary angle closure glaucoma in an east Asian population[224]. The intervention group were screened for risk of PACG. The control group were not screened. Screening failures were subjected to a more detailed ocular examination to determine suitability for laser iridotomy and baseline data regarding ocular status. The detailed ocular examination included visual acuity measured with the RLM-E chart. 770 subjects underwent visual acuity measurement using the same method as used in the Thai survey. Table A3 shows a sample of raw data from the Mongolian study.

Table A3. Sample raw data from the Mongolian glaucoma trial

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
31	number	r letters	l letters	r distance	l distance	r letters ph	l letters ph	r distance ph	l distance ph	r supra	l supra	r full	l full	RAPD
32	293	25	22	4	4	28	28	4	4	9	9	0	0	0
33	321	27	27	4	4	31	30	4	4	9	9	0	0	0
34	328	5	18	4	4	11	20	4	1	9	9	9	9	1
35	332	24	15	4	4	24	25	4	4	9	9	2	2	0
36	339	18	29	4	4	20	29	4	4	9	9	0	0	0
37	346	5	3	1	1	25	20	1	1	9	9	9	9	0
38	348	13	21	4	4	19	25	4	4	9	9	0	0	0
39	359	24	25	4	4	24	25	4	4	9	9	2	2	0
40	362	28	99	4	1	31	99	4	9	9	9	0	9	1
41	364	23	20	4	4	23	21	4	4	9	9	0	0	0
42	366	31	29	4	4	31	29	4	4	9	9	0	0	0
43	367	36	29	4	4	39	38	4	4	9	9	0	0	0
44	370	32	32	4	4	35	35	4	4	9	9	0	0	0
45	376	17	17	1	1	24	26	1	1	9	9	9	9	9
46	377	27	19	4	4	33	26	4	4	9	9	0	0	0
47	382	30	24	4	4	33	30	4	4	9	9	0	2	0
48	390	24	22	4	4	32	25	4	4	9	9	2	2	0
49	411	25	29	4	4	31	31	4	4	9	9	0	0	0
50	420	29	30	4	4	34	30	4	4	9	9	0	0	0
51	422	22	25	4	4	27	28	4	4	9	9	0	0	0
52	439	27	24	4	4	32	32	4	4	9	9	0	0	0
53	441	9	12	4	4	17	12	4	4	9	9	2	2	0
54	450	24	25	4	4	24	28	4	4	9	9	0	0	0
55	458	15	99	4	1	24	99	4	9	9	9	9	9	0
56	471	20	21	4	4	26	27	4	4	9	9	0	0	0
57	473	24	30	4	4	27	33	4	4	9	9	1	1	0
58	476	24	30	4	4	27	33	4	4	9	9	1	1	0
59	481	22	21	4	4	24	23	4	4	9	9	0	0	0
60	492	35	31	4	4	37	33	4	4	9	9	0	0	0

Appendix B

A statistical model to predict the sensitivity to change of visual acuity measurements (see section 16.2.1)

It is assumed that measurement errors are independent, identically, normally distributed random variables with mean 0 and variance σ^2.

Then the difference between two repeated measurements on same individual in the absence of any real change in acuity will be distributed as $N(0,2\sigma^2)$.

The choice of a change-criterion of ±1.96 standard deviations of the distribution of measured differences when the true change is 0 therefore corresponds to a change-criterion of $\pm 1.96 \times \sqrt{2}\sigma$. I.e. a measured change must be greater than $1.96 \times \sqrt{2}\sigma$ to be considered a real change.

Suppose two measurements $\hat{x}$ and $\hat{y}$ are made on an individual undergoing a true acuity change $d = x - y$.

Then the observed change $\hat{d} = \hat{x} - \hat{y} \sim N(d,2\sigma^2)$

The probability that the underlying change d is detected $= P\left(\left|N(d,2\sigma^2\right| > 1.96 \times \sqrt{2}\sigma\right)$.

Then, the probability that a deterioration d is detected as such

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \int_{1.96\sqrt{2}\sigma}^{\infty} e^{-\frac{(x-d)^2}{4\sigma^2}} dx$$

Let $d = a \times 1.96 \times \sqrt{2}\sigma$; i.e. re-express the change in terms of a multiples of the change-criterion.

Then the probability of detecting a deterioration of a times the change-criterion

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \int_{1.96\times\sqrt{2}\sigma}^{\infty} e^{\frac{(x-a\times 1.96\times\sqrt{2}\sigma)^2}{4\sigma^2}} dx$$

Let $y = x - a \times 1.96 \times \sqrt{2}\sigma$

Then the probability of detecting a deterioration of a times the change-criterion

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \int_{1.96\sqrt{2}\sigma(1-a)}^{\infty} e^{-\frac{y^2}{4\sigma^2}} dy$$

Let $\nu = \sqrt{2}\sigma$

Then the probability of detecting a deterioration of a times the change-criterion

$$= \frac{1}{\sqrt{2\pi}\nu} \int_{1.96\nu(1-a)}^{\infty} e^{-\frac{y^2}{2\nu^2}} dy$$

This is the probability that a normally distributed variable with mean 0 and variance ν^2 is greater than $1.96\nu(1-a)$ or that a standard normal variable (mean 0, variance 1) is greater than $1.96(1-a)$.

Appendix C

Glossary of terms and abbreviations

Term/ Abbreviation	Explanation
95% c.i.	95% confidence interval
95% TRR	95% test-retest range – the range over which 95% of differences between consecutive acuity scores will fall in the absence of any intervening change in clinical status (see thesis section 4.1.2)
Accuracy	This is the degree to which an estimate represents (on average) the true value of what is being measured (sections 2.7 & 4.1.1)
Agreement	The degree to which the results of one test can be substituted for those of another (see section 4.1.3)
Change-criterion	A cut-off by which measured change can be attributed to either a true change in clinical status, or TRV alone
CRT	Cathode ray tube – a traditional form of computer display
D	Dioptre(s) – a unit of focussing power (the reciprocal of the focal length of a lens in metres)
logMAR	Logarithm to the base 10 of the minimum angle of resolution
MAR	Minimum angle of resolution – the angular subtense of one element of the smallest resolvable optotype (in minutes of arc)
Precision	This represents the repeatability of a test. In statistical terms, it represents the inverse of the variance of a measurement or estimate (see section 4.1.2)
RLM	Reduced logMAR – a prototype visual acuity chart which features an abbreviated version of the Bailey-Lovie/ETDRS design
TFT	Thin film transistor – a form of flat-panel computer display
TRV	Test-retest variability – the variation in test score observed when a clinical test is repeated in the absence of any change in clinical status (can be thought of as measurement error or noise)

Appendix D

Peer-reviewed publications originating directly from this thesis

Rosser DA, Laidlaw DAH, Murdoch IE. The development of a 'Reduced logMAR' visual acuity chart for use in routine clinical practice. *Br J Ophthalmol* 2001;85:432-436

Rosser DA, Murdoch IE, Fitzke FW, Laidlaw DAH. Improving on ETDRS acuities: design and results for a computerised thresholding device. *Eye* 2003;17(6):701-6

Rosser DA, Murdoch IE, Cousens SN. The effect of optical defocus on the test-retest variability of visual acuity measurements. *Invest Ophthalmol Vis Sci* 2004;45:1076-1079

Rosser DA, Cousens SN, Murdoch IE, Fitzke FW, Laidlaw DAH. How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Invest Ophthalmol Vis Sci* 2003;44(8):3278-3281

Cousens SN, **Rosser DA**, Murdoch IE. A simple model to predict the sensitivity to change of visual acuity measurements. *Optom Vis Sci* 2004;81:673-677